O'Bryen, Barbara

From: Swope, Sheridan
Sent: Wednesday, December 31, 2003 1:40 PM
To: O'Bryen, Barbara
Subject: FW: ~~09980,818~~

Importance: High 09-980881

Scan this pg.

[Swope, Sheridan] -----Original Message-----
From: Swope, Sheridan
Sent: Monday, December 22, 2003 5:02 PM
To: STIC-Biotech/ChemLib
Subject: 09980,818

For 09980,818, pls search and interference search:

SID 1: Full-length and nt18-1097 against the NT and AA data bases
SID 2: against the NT and AA data bases
SID 3: against the NT and AA data bases
SID 4: against the NT and AA data bases
SID 9: against the NT and AA data bases.

THANKS!!!

Sheridan Swope, Ph.D.
Patent Examiner, AU 1652
Recombinant Enzymes
sheridan.swope@uspto.gov
703-305-1696 (voice)
703-308-3014 (FAX)
Mailbox: CM1 Rm10D01
Office: CM1 Rm12D12

As of January 12, 2003:
571-272-0943 (voice & FAX)
E03A70 Remsen Bld (Office)
E03A81 Remsen Bld (Mailbox)
400 Dulany Street
Alexandria, VA

Barb O'Bryen
01/08/04

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:38:01 ; Search time 206.037 Seconds
(without alignments)
2423.606 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135747_24935/app_query.fasta_1.2958
-DB=A_Geneseq_19Jun03 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_117_@runat_31122003_135747_24935 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1911	68.8	360	21	AAB11457	Human brain carbox
2	1907.5	68.7	386	24	ABG72095	Human protease pro
3	1902.5	68.5	386	24	ABG72093	Human protease pro
4	1885	67.9	423	17	AAR90293	Human plasma carbo
5	1885	67.9	423	18	AAW14733	Human plasma carbo
6	1885	67.9	423	20	AAW92270	Human plasma carbo
7	1880	67.7	423	14	AAR36273	Human plasma carbo
8	1785	64.3	338	21	AAB11458	Human brain carbox
9	1325	47.7	246	21	AAB11459	Human brain carbox
10	889	32.0	211	21	AAB58129	Lung cancer associ
11	652.5	23.5	417	24	ABB99688	Amino acid sequenc
12	651.5	23.5	415	17	AAW06172	Human pancreatic c
13	651.5	23.5	415	19	AAW74476	Human pancreatic c
14	647.5	23.3	417	23	AAU84325	Protein CPA3 diffe
15	647.5	23.3	667	24	ABP97414	Anti-human seminal
16	644	23.2	424	17	AAW06175	ProHCPB with PelB
17	640.5	23.1	404	16	AAR75131	Porcine Tyr-His-Me
18	636	22.9	324	22	AAG66562	Bothops jararaca c
19	635.5	22.9	437	22	AAE00502	Human carboxypepti
20	635.5	22.9	437	23	AAE19175	Human protease, PR
21	634.5	22.8	716	18	AAW41414	PreproHCPB-linker-
22	630.5	22.7	434	24	ABB99687	Amino acid sequenc
23	607.5	21.9	307	17	AAW00602	Rat carboxypeptida
24	576	20.7	306	16	AAR75132	Porcine carboxypep
25	576	20.7	306	22	AAU04477	Porcine carboxypep
26	575	20.7	349	17	AAW06173	mature HCPB-(His)6
27	572.5	20.6	329	17	AAW06174	Mature HCPB with P
28	571.5	20.6	613	18	AAW41412	Humanised Fd-mutan
29	571.5	20.6	613	19	AAW74481	Human pancreatic c
30	570	20.5	354	22	AAG66560	Human secreted met
31	570	20.5	354	22	AAG66566	Human secreted met
32	570	20.5	362	22	AAB74682	Human protease and
33	570	20.5	374	22	AAG66547	Human secreted met
34	570	20.5	374	22	AAG66565	Human secreted met
35	570	20.5	374	23	AAU82703	Amino acid sequenc
36	570	20.5	374	24	ABJ37888	NOVX protein seque
37	570	20.5	374	24	ABG74658	Human zinc-binding
38	568	20.5	349	17	AAW06181	Modified HCPB (D25
39	568	20.5	349	18	AAW13749	Carboxypeptidase B
40	567	20.4	349	17	AAW06182	Modified HCPB (D25
41	567	20.4	349	18	AAW13750	Carboxypeptidase B
42	564	20.3	349	18	AAW13763	Carboxypeptidase B
43	563	20.3	444	23	AAE24240	Human 23566 (carbo
44	562	20.2	349	18	AAW13752	Carboxypeptidase B
45	562	20.2	349	18	AAW13757	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAB11457
ID AAB11457 standard; Protein; 360 AA.
XX
AC AAB11457;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX

PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
DR N-PSDB; AAC81962.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 1; Page 68-71; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 360 AA;

Alignment Scores:

Pred. No.:	4.66e-197	Length:	360
Score:	1911.00	Matches:	360
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	68.82%	Indels:	0
DB:	21	Gaps:	0

US-09-980-881A-1 (1-1573) x AAB11457 (1-360)

```
Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220

Qy        678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1038 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
```

RESULT 2
ABG72095
ID ABG72095 standard; Protein; 386 AA.
XX
AC ABG72095;
XX
DT 14-FEB-2003 (first entry)
XX
DE Human protease protein, variation #2.
XX
KW Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 169
FT /note= "This residue is the minor variant of a single
FT nucleotide polymorphism. The major variant is an
FT Ala residue"
XX
PN US2002137179-A1.
XX
PD 26-SEP-2002.
XX
PF 21-MAR-2001; 2001US-0813133.
XX
PR 21-MAR-2001; 2001US-0813133.

```
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Page -; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #2, the gene for which is located on chromosome 13.
CC   Note: This sequence is not shown in the specification but has been
CC   created by the indexer from information given in figure 3.
XX
SQ   Sequence   386 AA;

Alignment Scores:
Pred. No.:              1.15e-196       Length:         386
Score:                  1907.50         Matches:        364
Percent Similarity:     94.56%          Conservative:   1
Best Local Similarity:  94.30%          Mismatches:     4
Query Match:            68.69%          Indels:         18
DB:                     24              Gaps:           1

US-09-980-881A-1 (1-1573) x ABG72095 (1-386)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy        678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977
              ||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1038 ATCAAATATTCGTT---------------------------------------------- 1051
              ||||||||||||
Db        341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360

Qy       1052 ------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGG 1105
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrp 380

Qy       1106 CATGTCATTAGGAATGTT 1123
              ||||||||||||||||||
Db        381 HisValIleArgAsnVal 386

RESULT 3
ABG72093
ID   ABG72093 standard; Protein; 386 AA.
XX
AC   ABG72093;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #1.
```

XX
KW Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 169
FT /note= "This residue is the major variant of a single
FT nucleotide polymorphism. The minor variant is an
FT Thr residue"
XX
PN US2002137179-A1.
XX
PD 26-SEP-2002.
XX
PF 21-MAR-2001; 2001US-0813133.
XX
PR 21-MAR-2001; 2001US-0813133.
XX
PA (GANW/) GAN W.
PA (KETC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Gan W, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR WPI; 2003-102387/09.
DR N-PSDB; ABX13670, ABX13671.
XX
PT New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT -
XX
PS Claim 1; Fig 2; 80pp; English.
XX
CC The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC protein, variant #1, the gene for which is located on chromosome 13.
XX
SQ Sequence 386 AA;

Alignment Scores:

Pred. No.:	3.99e-196	Length:	386
Score:	1902.50	Matches:	363
Percent Similarity:	94.30%	Conservative:	1
Best Local Similarity:	94.04%	Mismatches:	5
Query Match:	68.51%	Indels:	18

DB: 24 Gaps: 1

US-09-980-881A-1 (1-1573) x ABG72093 (1-386)

```
Qy      18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy      78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy     138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy     198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy     258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy     318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy     378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy     438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy     498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
           |||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db     161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy     558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy     618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy     678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737
           ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy     738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy     798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy     858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy     918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977
           ||||||||||||||||||||:::|||  ||||||||||||||||||||||||||||||||
Db     301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy     978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy    1038 ATCAAATATTCGTT-------------------------------------------- 1051
```

```
              ||||||||||||
Db        341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360

Qy       1052 ------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGG 1105
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrp 380

Qy       1106 CATGTCATTAGGAATGTT 1123
              ||||||||||||||||||
Db        381 HisValIleArgAsnVal 386

RESULT 4
AAR90293
ID   AAR90293 standard; Protein; 423 AA.
XX
AC   AAR90293;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active-site     181
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    183
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    256..257
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     310
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    312
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     369
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    370..371
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    394
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   N-PSDB; AAT11671.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

Alignment Scores:
Pred. No.:              3.24e-194      Length:         423
Score:                  1885.00        Matches:        365
Percent Similarity:     86.52%         Conservative:   1
Best Local Similarity:  86.29%         Mismatches:     3
Query Match:            67.88%         Indels:         55
DB:                     17             Gaps:           2

US-09-980-881A-1 (1-1573) x AAR90293 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 -----------------------------------------AATCGAATGTGGAGAAAG 626
                                                      ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
```

```
Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                      |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423
```

RESULT 5
AAW14733
ID AAW14733 standard; Protein; 423 AA.
XX
AC AAW14733;
XX
DT 25-MAR-2003 (updated)
DT 08-MAY-1997 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW plasma; plasminogen.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT		/note= "Signal peptide"
FT	Protein	23..423
FT		/note= "Mature PCPB"
FT	Cleavage-site	114
FT		/note= "Potential clip site for activation of PCPB
FT		as a carboxypeptidase"
FT	Active-site	181
FT		/note= "Expected to be involved in catalytic activity"
FT	Binding-site	183
FT		/note= "Expected to be involved in substrate binding"
FT	Active-site	184
FT		/note= "Expected to be involved in catalytic activity"
FT	Binding-site	256
FT		/note= "Expected to be involved in substrate binding"
FT	Binding-site	257
FT		/note= "Expected to be involved in substrate binding"
FT	Active-site	310
FT		/note= "Expected to be involved in catalytic activity"
FT	Binding-site	312
FT		/note= "Expected to be involved in substrate binding"
FT	Active-site	369
FT		/note= "Expected to be involved in catalytic activity"
FT	Binding-site	370
FT		/note= "Expected to be involved in substrate
FT		binding, thought to determine specificity
FT		of PCPB as a carboxypeptidase B"
FT	Binding-site	371
FT		/note= "Expected to be involved in substrate binding"
FT	Binding-site	394
FT		/note= "Expected to be involved in substrate binding"

XX
PN US5593674-A.
XX
PD 14-JAN-1997.
XX
PF 27-APR-1995; 95US-0430787.
XX
PR 01-FEB-1991; 91US-0649591.
PR 14-OCT-1992; 92US-0959944.
PR 15-DEC-1993; 93US-0167727.
PR 19-JUL-1994; 94US-0277540.
PR 27-APR-1995; 95US-0430787.
XX
PA (GETH) GENENTECH INC.
XX
PI Drayna DT, Eaton DL;
XX
DR WPI; 1997-099413/09.
DR N-PSDB; AAT62846.
XX
PT Using human plasma carboxy:peptidase B in blood coagulation - is
PT functionally related to carboxy:peptidase A and pancreas
PT carboxy:peptidase B
XX
PS Example 2; Column 37-42; 39pp; English.
XX
CC This sequence represents human plasma carboxypeptidase B (PCPB) which
CC has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC PCPB may be used therapeutically in haemostatic regulation. PCPB is
CC purified from human plasma or by transformed cell culture by
CC extraction using plasminogen bound to a solid phase.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 423 AA;

Alignment Scores:			
Pred. No.:	3.24e-194	Length:	423
Score:	1885.00	Matches:	365
Percent Similarity:	86.52%	Conservative:	1
Best Local Similarity:	86.29%	Mismatches:	3
Query Match:	67.88%	Indels:	55
DB:	18	Gaps:	2

US-09-980-881A-1 (1-1573) x AAW14733 (1-423)

```
Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
```

```
Qy       198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal  80

Qy       258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy       438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy       498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy       558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy       608 ------------------------------------------------------------ 608

Db       201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy       609 -----------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db       221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
            ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy      1047 TCGTT-------------------------------------------------------TAC 1054
            |||                                                         |||
Db       381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy      1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy      1115 AGGAATGTT 1123
```

```
            |||||||||
Db       421 ArgAsnVal 423
```

RESULT 6
AAW92270
ID AAW92270 standard; Protein; 423 AA.
XX
AC AAW92270;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT		/note= "signal peptide"
FT	Protein	23..423
FT		/note= "mature PCPBthr147"

XX
PN WO9855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI; 1999-045800/04.
DR N-PSDB; AAV74302.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Disclosure; Page 25-26; 35pp; English.
XX
CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 423 AA;

Alignment Scores:

Pred. No.:	3.24e-194	Length:	423
Score:	1885.00	Matches:	365
Percent Similarity:	86.52%	Conservative:	1
Best Local Similarity:	86.29%	Mismatches:	3
Query Match:	67.88%	Indels:	55
DB:	20	Gaps:	2

US-09-980-881A-1 (1-1573) x AAW92270 (1-423)

```
Qy        18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC  77
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
                |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ------------------------------------------AATCGAATGTGGAGAAAG 626
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
                ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
                |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
                |||||||||
Db        421 ArgAsnVal 423
```

RESULT 7
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX
AC AAR36273;
XX
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT Peptide   1..22
FT   /label= signal_peptide
FT Cleavage-site   114
FT   /note= "putative site of cleavage by trypsin to
FT   activate PCPB as a carboxypeptidase"
FT Active-site   181
FT   /note= "expected to be involved in catalytic
FT   activity"
FT Binding-site   183
FT   /note= "expected to be involved in substrate
FT   binding"
FT Active-site   184
FT   /note= "expected to be involved in catalytic
FT   activity"
FT Binding-site   256..257
FT   /note= "expected to be involved in substrate
FT   binding"
FT Active-site   310
FT   /note= "expected to be involved in catalytic
FT   activity"
FT Binding-site   312
FT   /note= "expected to be involved in substrate
FT   binding"
FT Active-site   369
FT   /note= "expected to be involved in catalytic
FT   activity"
FT Binding-site   370..371
FT   /note= "expected to be involved in substrate
FT   binding; Asp 370 is the residue which
FT   determines specificity of the PCPB as
FT   a carboxypeptidase B"
FT Binding-site   394
FT   /note= "expected to be involved in substrate
FT   binding"
XX
PN US5206161-A.
XX
PD 27-APR-1993.
XX
PF 01-FEB-1991; 91US-0649591.
XX
PR 01-FEB-1991; 91US-0649591.
XX
PA (GETH) GENENTECH INC.

```
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   N-PSDB; AAQ41001.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Claim 2; Fig 4; 40pp; English.
XX
CC   This amino acid sequence was deduced from a clone isolated from a
CC   human liver cDNA library. There is about 40% sequence identity
CC   between the deduced preproPCPB amino acid sequence and pro-human
CC   mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC   carboxypeptidase B. Human PCPB has the same substrate binding sites
CC   as, and shares 6 cysteine residues which form 3 intramolecular
CC   disulphide bonds with, bovine and rat carboxypeptidase B. The
CC   presence of the same amino acid (Asp at position 348 of the mature
CC   protein) at the region in carboxypeptidases that determines
CC   substrate specificity, suggests that PCPB represents a plasma-
CC   derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC   tPA of plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence   423 AA;

Alignment Scores:
Pred. No.:                1.12e-193      Length:          423
Score:                    1880.00        Matches:         364
Percent Similarity:       86.29%         Conservative:    1
Best Local Similarity:    86.05%         Mismatches:      4
Query Match:              67.70%         Indels:          55
DB:                       14             Gaps:            2

US-09-980-881A-1 (1-1573) x AAR36273 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ------------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaCysSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT-----------------------------------------------------TAC 1054
              |||                                                     |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 8
AAB11458
ID   AAB11458 standard; Protein; 338 AA.
XX
AC   AAB11458;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
```

```
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 71-73; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   338 AA;

Alignment Scores:
Pred. No.:              1.82e-183      Length:         338
Score:                  1785.00        Matches:        335
Percent Similarity:     99.41%         Conservative:   1
Best Local Similarity:  99.11%         Mismatches:     2
Query Match:            64.28%         Indels:         0
DB:                     21             Gaps:           0

US-09-980-881A-1 (1-1573) x AAB11458 (1-338)

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Qy        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Qy        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Qy        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Qy        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Qy        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 623
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180

Qy        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200

Qy        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220

Qy        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240

Qy        804 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 863
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProPro 260

Qy        864 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 923
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280

Qy        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300

Qy        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1043
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320

Qy       1044 TATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 TyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338

RESULT 9
AAB11459
ID   AAB11459 standard; Protein; 246 AA.
XX
AC   AAB11459;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 74-75; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
```

```
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   246 AA;

Alignment Scores:
Pred. No.:              7.22e-134       Length:         246
Score:                  1325.00         Matches:        243
Percent Similarity:     99.19%          Conservative:   1
Best Local Similarity:  98.78%          Mismatches:     2
Query Match:            47.71%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-881A-1 (1-1573) x AAB11459 (1-246)

Qy        360 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

Qy        420 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

Qy        480 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 539
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60

Qy        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80

Qy        600 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 659
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100

Qy        660 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 719
              |||||||||||||||  ||||||  ||||||||||||||||||||||||||||||||||||
Db        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120

Qy        720 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 779
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140

Qy        780 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160

Qy        840 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 899
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180

Qy        900 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959
              ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200

Qy        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220

Qy       1020 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTT 1079
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TrpIleTyrAspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeu 240

Qy       1080 TTGCCGCTGTCTCTAAAA 1097
              ||||||||||||||||||
Db        241 LeuProLeuSerLeuLys 246
```

```
RESULT 10
AAB58129
ID   AAB58129 standard; Protein; 211 AA.
XX
AC   AAB58129;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18005.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 944-945; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   211 AA;

Alignment Scores:
Pred. No.:              7.96e-87        Length:         211
Score:                  889.00          Matches:        173
Percent Similarity:     83.33%          Conservative:   2
Best Local Similarity:  82.38%          Mismatches:     9
Query Match:            32.01%          Indels:         27
DB:                     21              Gaps:           3

US-09-980-881A-1 (1-1573) x AAB58129 (1-211)

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------------GGC 605
              |||||||||          |||   |||   :::
```

```
Db          3 TrpIleSer---MetLeuCysArgTrpLeuMetTrpMetValMetAsnTyrSerTrpLys 21

Qy        606 CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACA 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         22 LysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThr 41

Qy        666 GACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCA 725
              |||||||||   ||||||   ||||||||||||||||||||||||||||||||||||||
Db         42 AspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSer 61

Qy        726 TGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCT 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAla 81

Qy        786 AGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 SerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSer 101

Qy        846 CAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 GlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 121

Qy        906 TCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTAT 965
              ||||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db        122 SerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyr 141

Qy        966 ACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATC 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 ThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIle 161

Qy       1026 TATGATTTGGGCATCAAATATTCGTT---------------------------------- 1051
              ||||||||||||||||||||||||
Db        162 TyrAspLeuGlyIleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGly 181

Qy       1052 ------------------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCT 1093
                               ||||||||||||||||||||||||||||||||||||||||||
Db        182 PheLeuLeuProGluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSer 201

Qy       1094 AAAATAGCTTGGCATGTCATTAGGAATGTT 1123
              |||||||||||||||||||||||||||||
Db        202 LysIleAlaTrpHisValIleArgAsnVal 211

RESULT 11
ABB99688
ID   ABB99688 standard; Protein; 417 AA.
XX
AC   ABB99688;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Amino acid sequence of human zinc carboxypeptidase B precursor.
XX
KW   Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW   chronic obstructive pulmonary disease; COPD; haematological disorder;
KW   zinc carboxypeptidase B.
XX
OS   Homo sapiens.
XX
PN   WO200295018-A1.
XX
PD   28-NOV-2002.
XX
PF   16-MAY-2002; 2002WO-EP05392.
XX
PR   22-MAY-2001; 2001US-292329P.
XX
PA   (FARB ) BAYER AG.
XX
PI   Smolyar A;
XX
DR   WPI; 2003-120801/11.
XX
PT   New polynucleotide encoding a zinc carboxypeptidase B-like protein
PT   polypeptide, useful for treating diseases associated with
PT   carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT   haematological disorder  -
XX
PS   Disclosure; Fig 3; 113pp; English.
XX
CC   The present sequence represents a zinc carboxypeptidase B precursor.
CC   The specification describes a human zinc carboxypeptidase B-like
CC   protein. The zinc carboxypeptidase B-like protein is used for treating
CC   diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
CC   disease (COPD) or haematological disorders. The zinc carboxypeptidase
CC   B-like protein may also be used to identify compounds which may act
CC   as activators or inhibitors at the enzyme's active site, to raise
CC   specific antibodies which can block the enzyme and effectively reduce
CC   its activity, as a bait protein in a two-hybrid or three-hybrid assay
CC   to identify other proteins which bind to or interact with the human
CC   zinc carboxypeptidase-like protein polypeptide and modulate its
CC   activity, and for immunization of mammals.
XX
SQ   Sequence   417 AA;

Alignment Scores:
Pred. No.:              3.81e-61        Length:         417
Score:                  652.50          Matches:        142
Percent Similarity:     52.13%          Conservative:   78
Best Local Similarity:  33.65%          Mismatches:     141
Query Match:            23.50%          Indels:         62
DB:                     24              Gaps:           8

US-09-980-881A-1 (1-1573) x ABB99688 (1-417)

Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
              :::   :::|||   |||||||||   :::   |||           |||
Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         84 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
                 |||:::      :::|||                                :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy        141 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 200
              ::::::   |||   :::|||   :::|||      |||:::|||   :::
Db         41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        201 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 260
              |||           |||   |||   |||   |||   |||           |||:::
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy        261 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 320
                 |||      :::   :::         ||||||:::::::::::::   ::::::
Db         81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy        321 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 380
              ::::::   |||      :::                 ||||||:::         ||||||
Db        101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
              :::|||:::          |||   :::|||   :::   :::         :::|||
Db        118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
                 ::::::::::::      ||||||::::::||||||         :::|||:::|||
Db        138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
              ||||||   ||||||            |||   ||||||::::::|||||||||   |||
Db        158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 602
              ||||||||||||||||||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy        602 ------------------------------------------------------------ 602
```

```
Db        197 TyrGlyArgGluIleGlnValThrGluLeuLeuAsnLysLeuAspPheTyrValLeuPro 216

Qy        603 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629
                             |||:::                 :::|||   |||||||||
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy        630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
              ||||||   :::      :::   ||||||||||||||   |||   ||||||
Db        237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
                |||||||||   |||||||||   :::   |||   ||||||||||||||
Db        256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
                |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
                ||||||||||:::::::::|||||||||||||||   :::::::::|||||||||
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
              |||                 :::::::::   ||||||:::   :::|||         |||
Db        316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
              :::   :::      :::          |||:::||||||:::|||   |||:::   |||
Db        336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
              :::|||   |||   ||||||   ||||||||   ||||||   ||||||:::||||||
Db        355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

Qy       1050 TTTACATC---------------------------------------------------- 1057
              ||||||
Db        375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

Qy       1058 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGG 1117
              :::   ||||||   |||   |||   |||:::      :::|||   :::|||:::
Db        395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 414

Qy       1118 AATGTT 1123
              ::::::
Db        415 HisLeu 416
```

RESULT 12
AAW06172
ID AAW06172 standard; Protein; 415 AA.
XX
AC AAW06172;
XX
DT 12-FEB-1997 (first entry)
XX
DE Human pancreatic carboxypeptidase B.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..13
FT /note= "partial enzyme secretion leader sequence;
FT (pre-sequence)"
FT Peptide 14..108
FT /label= pro-sequence
FT Protein 109..415
FT /label= mature_enzyme
FT Misc-difference 243
FT /note= "not observed in other published HPCB
FT sequences"
FT Active-site 361
FT /note= "Asp253 characteristic of B type specificity"
XX
PN WO9620011-A1.
XX
PD 04-JUL-1996.
XX
PF 21-DEC-1995; 95WO-GB02991.
XX
PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
PA (ZENE) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI; 1996-321650/32.
DR N-PSDB; AAT42494.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti:neoplastic prodrug
XX
PS Reference Example 14; Page 131-132; 182pp; English.
XX
CC A two-component system for anti-tumour therapy comprises a targeting
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC prodrug. The system is based on antibody directed enzyme prodrug therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence is that of the native prepro-HCPB.
XX
SQ Sequence 415 AA;

Alignment Scores:

Pred. No.:	4.87e-61	Length:	415
Score:	651.50	Matches:	142
Percent Similarity:	52.03%	Conservative:	76
Best Local Similarity:	33.89%	Mismatches:	140
Query Match:	23.46%	Indels:	62
DB:	17	Gaps:	8

US-09-980-881A-1 (1-1573) x AAW06172 (1-415)

```
Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 89
              |||   |||||||||   :::   |||            |||            |||:::
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         90 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 149
                 :::|||                          :::   :::::::::
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleIleArgGlu 41

Qy        150 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 209
              |||   :::|||   :::|||      |||:::|||   :::         |||
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        210 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 269
                  |||   |||   |||   |||   |||         |||:::      |||
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        270 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 329
                    :::   :::          |||||:::::::::::   :::::::::::
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        330 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
              |||      :::               ||||||:::         ||||||:::|||:::
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118
```

```
Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 602
Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        602 ------------------------------------------------------------ 602

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        603 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 638
Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1057
Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1058 ----------------------------------------------------AAACCCACC 1066
Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123
Db        396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414

RESULT 13
AAW74476
ID   AAW74476 standard; Protein; 415 AA.
XX
AC   AAW74476;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Human pancreatic carboxypeptidase protein sequence.
XX
KW   Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW   prodrug therapy.
XX
OS   Homo sapiens.
XX
PN   WO9835988-A1.
XX
PD   20-AUG-1998.
XX
PF   10-FEB-1998;   98WO-GB00415.
XX
PR   29-OCT-1997;   97GB-0022727.
PR   14-FEB-1997;   97GB-0003104.
PR   18-OCT-1997;   97GB-0022003.
XX
PA   (ZENE ) ZENEPA   (ZENE ) ZENECA LTD.
XX
PI   Edge MD;
XX
DR   WPI; 1998-467168/40.
DR   N-PSDB; AAV41795.
XX
PT   New modified pro-domain of carboxy-peptidase B - enhances expression
PT   of co-expressed proteins for production of recombinant
PT   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT   enzyme prodrug therapy
XX
PS   Example 1; Page 54-55; 83pp; English.
XX
CC   The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC   modified pro-domain of CPB on a separate gene to enhance recombinant
CC   expression.  This process can be used to produce recombinant CPB in
CC   eukaryotic cells, or fusions of CPB with antibody chains.  CPB is used in
CC   insulin production and protein sequencing, while its fusions with
CC   antibodies are useful in antibody-directed enzyme prodrug therapy.  The
CC   modified pro-domain provide increased yields of recombinant CPB, possibly
CC   by protecting the C-terminus against enzymatic degradation or by
CC   increasing intracellular trafficking.
XX
SQ   Sequence   415 AA;

Alignment Scores:
Pred. No.:              4.87e-61        Length:         415
Score:                  651.50          Matches:        142
Percent Similarity:     52.03%          Conservative:   76
Best Local Similarity:  33.89%          Mismatches:     140
Query Match:            23.46%          Indels:         62
DB:                     19              Gaps:           8

US-09-980-881A-1 (1-1573) x AAW74476 (1-415)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 89
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         90 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 149
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        150 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 209
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        210 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 269
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        270 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 329
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        330 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
```

```
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509

Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569

Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 602

Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        602 ------------------------------------------------------------ 602

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        603 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 638

Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698

Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758

Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818

Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878

Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938

Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998

Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1057

Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1058 ----------------------------------------------------AAACCCACC 1066

Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123

Db        396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414

RESULT 14
AAU84325
ID   AAU84325 standard; Protein; 417 AA.
XX
AC   AAU84325;
XX
DT   08-MAY-2002  (first entry)
XX
DE   Protein CPA3 differentially expressed in breast cancer tissue.
XX
KW   Human; diagnosis of breast cancer; endometrial cancer; breast tumour;
KW   MAI; mitotic activity index; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO200210436-A2.
XX
PD   07-FEB-2002.
XX
PF   27-JUL-2001; 2001WO-US23642.
XX
PR   28-JUL-2000; 2000US-222093P.
XX
PA   (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.
PA   (BAAK/) BAAK J.
XX
PI   Baak J,  Mutter GL;
XX
DR   WPI; 2002-180084/23.
DR   N-PSDB; ABK35545.
XX
PT   Diagnosing breast cancer comprises determining expression of nucleic
PT   acid molecules or expression products that are differentially expressed
PT   in normal and malignant tissue -
XX
PS   Claim 37; Page 134-135; 219pp; English.
XX
CC   The present invention relates to a method for diagnosing breast cancer
CC   in a subject suspected of having endometrial cancer. The method
CC   comprises determining the expression of a set of human genes or
CC   expression products in an endometrial sample suspected of being
CC   cancerous. The human genes of the invention are differentially
CC   expressed in breast tumours characterised as high or low MAI (mitotic
CC   activity index). These sets of genes can be used to discriminate between
CC   high and low MAI breast tumours. The invention also provides DNA and
CC   protein microarrays for analysing the expression of the human genes and
CC   their protein products. The methods and arrays are useful for the
CC   diagnosis and prognosis of endometrial cancer, selecting and monitoring
CC   treatment regimes, and identification of compounds useful for the
CC   treatment of endometrial cancer. AAU84311-AAU84361 represent the human
CC   proteins of the invention that are differentially expressed in breast
CC   cancer tissue.
XX
SQ   Sequence   417 AA;

Alignment Scores:
Pred. No.:              1.32e-60      Length:          417
Score:                  647.50        Matches:         143
Percent Similarity:     52.46%        Conservative:    81
Best Local Similarity:  33.49%        Mismatches:      132
Query Match:            23.32%        Indels:          72
DB:                     23            Gaps:            10

US-09-980-881A-1 (1-1573) x AAU84325 (1-417)

Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65

Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125

Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185

Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245

Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305

Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
```

```
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              |||         |||   :::|||::::::      :::|||   :::|||   |||
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              :::   :::::::::|||:::|||::::::::::|||   |||||||||::   |||
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
              ||||||||||||||||||:::   |||::::::         :::   ||||||:::
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
              :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
                                             |||:::                 |||
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              ||||||||||||||||||||||         |||:::   ||||||||||||||||||
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   ||||||   :::      |||               ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||     ||||||||||   |||   ||||||||  :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::|||||||||   ||||||               :::||||||:::|||:::
Db        310 LeuLeuProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                :::      |||      |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
              |||           |||:::|||         |||         ||||||   ||||||
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
              ||||||||||||:::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                           ||||||||||||||||||||         ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 15
ABP97414
ID   ABP97414 standard; Protein; 667 AA.
XX
AC   ABP97414;
XX
DT   30-MAY-2003  (first entry)
XX
DE   Anti-human seminal plasma protein ScFv/hCPA fusion protein.
XX
KW   Single chain antibody; ScFv; anti-human seminal plasma protein;
KW   monoclonal antibody E4B7; mouse; murine; human; carboxypeptidase A;
KW   hCPA; fusion protein; antibody-directed enzyme prodrug therapy; ADEPT;
KW   methotrexate-alpha-peptide prodrug; prostate cancer; cytostatic;
KW   gene; ds.
XX
OS   Chimeric - Mus musculus.
OS   Chimeric - Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Protein         1..245
FT                   /label= E4B7_derived_ScFv
FT                   /note= "Anti-human seminal plasma protein single
FT                         chain antibody derived from murine monoclonal
FT                         antibody E4B7 VH and VL regions; targets
FT                         fusion protein to prostate cancer cells"
FT   Region          1..122
FT                   /note= "Corresponds to monoclonal antibody E4B7 VH (heavy
FT                         chain variable region"
FT   Region          123..137
FT                   /label= ScFv_linker
FT   Region          138..245
FT                   /note= "Corresponds to monoclonal antibody E4B7 VL (light
FT                         chain variable region"
FT   Misc-difference 216
FT                   /note= "Encoded by GAC"
FT   Region          245..251
FT                   /label= ScFv/hCPA_linker
FT   Misc-difference 247
FT                   /note= "Encoded by AGC"
FT   Protein         253..667
FT                   /label= hCPA
FT                   /note= "Human carboxypeptidase A; activates methotrexate-
FT                         alpha-peptide prodrug"
XX
PN   WO2002100431-A1.
XX
PD   19-DEC-2002.
XX
PF   08-JUN-2001; 2001WO-CN00924.
XX
PR   08-JUN-2001; 2001WO-CN00924.
XX
PA   (UYTA-) UNIV TANGDU HOSPITAL FORTH MILITARY MEDI.
PA   (HAOX/) HAO X.
XX
PI   Hao X;
XX
DR   WPI; 2003-156924/15.
DR   N-PSDB; ABZ75116.
XX
PT   Pharmaceutical kits for use in the treatment of prostate cancer,
PT   comprise anti-human seminal plasma protein single-chain antibody/human
PT   carboxypeptidase fusion protein and a prodrug of
PT   methotrexate-alpha-peptides -
XX
PS   Claim 3; Fig 4; 29pp; Chinese.
XX
CC   The invention relates to a pharmaceutical kit for antibody-directed
CC   enzyme prodrug therapy (ADEPT) to treat prostate cancer. The kit
CC   comprises several respective containers for holding anti-human seminal
CC   plasma protein single-chain antibody (ScFv)/human carboxypeptidase A
CC   (hCPA) fusion protein (ABP97414), a prodrug of methotrexate-alpha-peptide
CC   and a pharmaceutically-acceptable auxiliary for medication. The invention
CC   also encompasses the anti-human seminal plasma protein ScFv/hCPA fusion
CC   protein, and the nucleic acid encoding it (ABZ75116). This was
CC   constructed by linking DNA encoding the ScFv (itself generated by linking
CC   DNAs encoding the heavy and light chain variable regions (VH and VL) of
CC   the murine anti-human seminal plasma protein monoclonal antibody E4B7)
```

CC to DNA encoding human carboxypeptidase A via a linker. The ScFv portion
CC of the fusion protein enables it to be specifically targeted to prostate
CC cancer cells, where the hCPA portion can then activate the prodrug which
CC in turn has cytostatic activity towards the cancer cells. The invention
CC permits the fusion protein to be easily produced in large quantities for
CC use in ADEPT for prostate cancer treatment. The present sequence
CC represents the ScFv/hCPA fusion protein of the invention.
XX
SQ Sequence 667 AA;

Alignment Scores:

Pred. No.:	1.66e-60	Length:	667
Score:	647.50	Matches:	143
Percent Similarity:	52.46%	Conservative:	81
Best Local Similarity:	33.49%	Mismatches:	132
Query Match:	23.32%	Indels:	72
DB:	24	Gaps:	10

US-09-980-881A-1 (1-1573) x ABP97414 (1-667)

```
Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
              :::   ||||||:::         :::||||||:::      |||:::
Db        255 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 269

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                          |||       :::|||           |||:::
Db        270 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 284

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
              :::|||      ::::::::::::|||      |||   |||:::      |||   |||
Db        285 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 304

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                        :::                 |||   |||   |||:::      :::
Db        305 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 324

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                  :::::::::   |||:::   ::.   :::         :::|||:::
Db        325 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 344

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
              |||:::::::::   |||::::::|||            :::   :::      |||
Db        345 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 364

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              |||         |||   :::|||::::::      :::|||   :::|||   |||
Db        365 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 381

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              :::   :::::::::|||:::|||:::::::::|||   |||||||||:::   |||
Db        382 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 401

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                 ||||||||||||||||||:::   |||::::::          :::   ||||||:::
Db        402 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 420

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
              :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        421 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 440

Qy        602 ------------------------------------------------------------ 602

Db        441 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 460

Qy        603 --------------------------------GGCCAT----------------AAT 611
                                             |||:::                |||
Db        461 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 480

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              ||||||||||||||||||||||         |||:::   ||||||||||||||||||
Db        481 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 500

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
```

```
                  |||   ||||||   :::      |||              ::::::      |||:::
Db        501 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 519

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||      ||||||||||||   |||   |||||||||   :::|||
Db        520 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 539

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||||
Db        540 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 559

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::|||||||||   ||||||               :::||||||:::|||:::
Db        560 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 579

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                  :::      |||          |||||||||   :::
Db        580 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 598

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
              |||             |||:::|||        |||         ||||||   ||||||
Db        599 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 618

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
              ||||||||||||:::::|||
Db        619 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 638

Qy       1055 ---------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                            ||||||||||||||||||||      ||||||      |||
Db        639 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 658

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::::
Db        659 AlaLysTyrIleLeuLysHis 665
```

Search completed: January 1, 2004, 18:52:39
Job time : 231.037 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:40:51 ; Search time 42.3933 Seconds
(without alignments)
3139.880 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135750_24983/app_query.fasta_1.2958
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_31_@runat_31122003_135750_24983 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1902.5	68.5	386	4	US-09-813-133A-2	Sequence 2, Appli
2	1885	67.9	423	1	US-07-649-591B-3	Sequence 3, Appli
3	1885	67.9	423	1	US-08-277-540-3	Sequence 3, Appli
4	1885	67.9	423	1	US-08-430-787A-3	Sequence 3, Appli
5	1885	67.9	423	2	US-08-869-057-2	Sequence 2, Appli
6	1874	67.5	423	4	US-09-813-133A-4	Sequence 4, Appli
7	655	23.6	417	1	US-07-649-591B-7	Sequence 7, Appli
8	655	23.6	417	1	US-08-277-540-7	Sequence 7, Appli
9	655	23.6	417	1	US-08-430-787A-7	Sequence 7, Appli
10	651.5	23.5	415	2	US-08-860-882A-57	Sequence 57, Appl
11	651.5	23.5	415	4	US-09-011-769A-39	Sequence 39, Appl
12	647.5	23.3	417	1	US-07-649-591B-6	Sequence 6, Appli
13	647.5	23.3	417	1	US-08-277-540-6	Sequence 6, Appli
14	647.5	23.3	417	1	US-08-430-787A-6	Sequence 6, Appli
15	644	23.2	424	4	US-09-011-769A-56	Sequence 56, Appl
16	640.5	23.1	404	1	US-08-696-139-2	Sequence 2, Appli
17	635.5	22.9	437	4	US-09-675-305-10	Sequence 10, Appl
18	634.5	22.8	716	3	US-09-171-945-125	Sequence 125, App
19	607.5	21.9	307	2	US-08-782-760-6	Sequence 6, Appli
20	607.5	21.9	307	5	PCT-US96-00995-6	Sequence 6, Appli
21	607	21.9	396	1	US-07-649-591B-4	Sequence 4, Appli
22	607	21.9	396	1	US-08-277-540-4	Sequence 4, Appli
23	607	21.9	396	1	US-08-430-787A-4	Sequence 4, Appli
24	576	20.7	306	1	US-08-696-139-4	Sequence 4, Appli
25	575	20.7	349	4	US-09-011-769A-47	Sequence 47, Appl
26	572.5	20.6	329	4	US-09-011-769A-51	Sequence 51, Appl
27	571.5	20.6	613	3	US-09-171-945-113	Sequence 113, App
28	568	20.5	349	4	US-09-011-769A-60	Sequence 60, Appl
29	567	20.4	349	4	US-09-011-769A-64	Sequence 64, Appl
30	553	19.9	417	1	US-07-649-591B-8	Sequence 8, Appli
31	553	19.9	417	1	US-08-277-540-8	Sequence 8, Appli
32	553	19.9	417	1	US-08-430-787A-8	Sequence 8, Appli
33	546.5	19.7	399	4	US-09-710-099-8	Sequence 8, Appli
34	529.5	19.1	350	4	US-09-675-305-12	Sequence 12, Appl
35	529	19.0	419	1	US-07-649-591B-5	Sequence 5, Appli
36	529	19.0	419	1	US-08-277-540-5	Sequence 5, Appli
37	529	19.0	419	1	US-08-430-787A-5	Sequence 5, Appli
38	520	18.7	436	4	US-09-710-099-6	Sequence 6, Appli
39	513	18.5	417	3	US-08-640-906-4	Sequence 4, Appli
40	513	18.5	417	4	US-09-395-936-4	Sequence 4, Appli
41	512	18.4	417	3	US-08-640-906-18	Sequence 18, Appl
42	512	18.4	417	4	US-09-395-936-18	Sequence 18, Appl
43	487.5	17.6	419	3	US-08-640-906-2	Sequence 2, Appli
44	487.5	17.6	419	4	US-09-395-936-2	Sequence 2, Appli
45	484	17.4	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Alignment Scores:

Pred. No.:	1.4e-202	Length:	386
Score:	1902.50	Matches:	363
Percent Similarity:	94.30%	Conservative:	1
Best Local Similarity:	94.04%	Mismatches:	5
Query Match:	68.51%	Indels:	18
DB:	4	Gaps:	1

US-09-980-881A-1 (1-1573) x US-09-813-133A-2 (1-386)

```
Qy          18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy          78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy       138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy       198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy       258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy       438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy       498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
             ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy       558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy       618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy       678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737
             ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy       738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy       798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy       858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy       918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977
             ||||||||||||||||||||:::|||  ||||||||||||||||||||||||||||||||
Db       301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy       978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy      1038 ATCAAATATTCGTT---------------------------------------------- 1051
             ||||||||||||
Db       341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360

Qy      1052 ------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGG 1105
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrp 380

Qy      1106 CATGTCATTAGGAATGTT 1123
             ||||||||||||||||||
Db       381 HisValIleArgAsnVal 386
```

```
RESULT 2
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-3

Alignment Scores:
Pred. No.:              1.29e-200      Length:          423
Score:                  1885.00        Matches:         365
Percent Similarity:     86.52%         Conservative:    1
Best Local Similarity:  86.29%         Mismatches:      3
Query Match:            67.88%         Indels:          55
DB:                     1              Gaps:            2

US-09-980-881A-1 (1-1573) x US-07-649-591B-3 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
```

```
Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-3

Alignment Scores:
Pred. No.:              1.29e-200      Length:          423
Score:                  1885.00        Matches:         365
Percent Similarity:     86.52%         Conservative:    1
Best Local Similarity:  86.29%         Mismatches:      3
Query Match:            67.88%         Indels:          55
DB:                     1              Gaps:            2

US-09-980-881A-1 (1-1573) x US-08-277-540-3 (1-423)

Qy          18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy          78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy         138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy         198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy         258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
```

```
Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 4
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-3

Alignment Scores:
Pred. No.:              1.29e-200       Length:         423
Score:                  1885.00         Matches:        365
Percent Similarity:     86.52%          Conservative:   1
Best Local Similarity:  86.29%          Mismatches:     3
Query Match:            67.88%          Indels:         55
DB:                     1               Gaps:           2

US-09-980-881A-1 (1-1573) x US-08-430-787A-3 (1-423)

Qy          18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy          78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy         138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy         198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy         258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy       438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy       498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy       558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy       608 ------------------------------------------------------------ 608

Db       201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy       609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db       221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
             ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy      1047 TCGTT-------------------------------------------------------TAC 1054
             |||                                                         |||
Db       381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy      1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy      1115 AGGAATGTT 1123
             |||||||||
Db       421 ArgAsnVal 423

RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057

; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  Plasma
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  23..401
US-08-869-057-2

Alignment Scores:
Pred. No.:                1.29e-200      Length:          423
Score:                    1885.00        Matches:         365
Percent Similarity:       86.52%         Conservative:    1
Best Local Similarity:    86.29%         Mismatches:      3
Query Match:              67.88%         Indels:          55
DB:                       2              Gaps:            2

US-09-980-881A-1 (1-1573) x US-08-869-057-2 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
```

```
Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423
```

RESULT 6
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294

; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-4

Alignment Scores:

Pred. No.:	2.14e-199	Length:	423
Score:	1874.00	Matches:	363
Percent Similarity:	86.05%	Conservative:	1
Best Local Similarity:	85.82%	Mismatches:	5
Query Match:	67.48%	Indels:	55
DB:	4	Gaps:	2

US-09-980-881A-1 (1-1573) x US-09-813-133A-4 (1-423)

```
Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240
```

```
Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
Db       241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
Db       261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
Db       281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
Db       301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db       321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
Db       341 ValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
Db       361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy      1047 TCGTT----------------------------------------------------TAC 1054
Db       381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy      1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
Db       401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy      1115 AGGAATGTT 1123
Db       421 ArgAsnVal 423
```

```
RESULT 7
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-7

Alignment Scores:
Pred. No.:              6.12e-64        Length:         417
Score:                  655.00          Matches:        150
Percent Similarity:     51.67%          Conservative:   66
Best Local Similarity:  35.89%          Mismatches:     138
Query Match:            23.59%          Indels:         65
DB:                     1               Gaps:           9

US-09-980-881A-1 (1-1573) x US-07-649-591B-7 (1-417)

Qy        33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 92
Db         7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy        93 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 152
Db        24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy       153 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 212
Db        44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy       213 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 272
Db        64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy       273 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 332
Db        84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy       333 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
Db       104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy       390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
Db       121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy       450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
Db       141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy       510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
Db       160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy       570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 608
Db       180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy       609 AAT---------------------------------------------------------- 611
Db       200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy       612 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 638
Db       220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy       639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
Db       240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy       699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
```

```
                          ::::::      |||         |||   |||     ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
              ||||||   |||   |||||||||   :::|||:::|||   ::::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
              |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
                      :::|||::::::|||      ||||||      |||      |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
                  |||         |||||||||   :::|||      :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1054
                  |||         ||||||:::||||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1055 ------------------------------------------------ATCAAACCCACC 1066
                                                              ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
              |||:::|||         :::|||      ||||||   ::::::::::::|||
Db        398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415

RESULT 8
US-08-277-540-7
; Sequence 7, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:

;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-7

Alignment Scores:
Pred. No.:              6.12e-64        Length:         417
Score:                  655.00          Matches:        150
Percent Similarity:     51.67%          Conservative:   66
Best Local Similarity:  35.89%          Mismatches:     138
Query Match:            23.59%          Indels:         65
DB:                     1               Gaps:           9

US-09-980-881A-1 (1-1573) x US-08-277-540-7 (1-417)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 92
              :::||||||:::       :::           |||       |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         93 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 152
                 :::|||              :::      :::         ||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        153 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 212
              |||   :::   |||:::      |||   |||                 |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        213 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 272
                 |||   |||   |||:::      :::         :::::::::   |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        273 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 332
                    |||         :::|||:::   |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        333 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
                       |||   ::::::   |||   |||         |||   :::|||:::
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
                  ::::::|||   ||||||   |||   :::   |||:::||||||:::|||:::
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
              ::::::|||   |||||||||:::   |||       |||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
              ||||||:::         :::   ||||||::::::||||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 608
              ||||||||||||||||||   ||||||:::                      |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        609 AAT---------------------------------------------------------- 611
              |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        612 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 638
                                             ||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
                 |||:::   ||||||||||||||||||||   ||||||              |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
                          ::::::      |||         |||   |||     ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
```

```
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1054
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1055 ----------------------------------------------ATCAAACCCACC 1066
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
Db        398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415

RESULT 9
US-08-430-787A-7
; Sequence 7, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-7

Alignment Scores:
Pred. No.:              6.12e-64      Length:          417
Score:                  655.00        Matches:         150
Percent Similarity:     51.67%        Conservative:    66
Best Local Similarity:  35.89%        Mismatches:      138
Query Match:            23.59%        Indels:          65
DB:                     1             Gaps:            9

US-09-980-881A-1 (1-1573) x US-08-430-787A-7 (1-417)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 92
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         93 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 152
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        153 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 212
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        213 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 272
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        273 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 332
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        333 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle-- 159

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------GGCCAT 608
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        609 AAT---------------------------------------------------------- 611
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        612 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 638
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298
```

```
Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
              |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
                     :::|||::::::|||      ||||||      |||      |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
                  |||         |||||||||   :::|||      :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1054
                  |||         ||||||:::||||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1055 ------------------------------------------------ATCAAACCCACC 1066
                                                              ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
              |||:::|||         :::|||      ||||||   ::::::::::::|||
Db        398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415

RESULT 10
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  415 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-860-882A-57
```

```
Alignment Scores:
Pred. No.:              1.5e-63        Length:          415
Score:                  651.50         Matches:         142
Percent Similarity:     52.03%         Conservative:    76
Best Local Similarity:  33.89%         Mismatches:      140
Query Match:            23.46%         Indels:          62
DB:                     2              Gaps:            8

US-09-980-881A-1 (1-1573) x US-08-860-882A-57 (1-415)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 89
              |||   |||||||||   :::   |||           |||           |||:::
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         90 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 149
                 :::|||                                   :::   :::::::::
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        150 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 209
              |||   :::|||   :::|||      |||:::|||   :::         |||
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        210 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 269
                    |||   |||   |||   |||   |||         |||:::      |||
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        270 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 329
                  :::   :::          |||||:::::::::::   ::::::::::::
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        330 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
              |||      :::              ||||||:::         ||||||:::|||:::
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
                    |||   :::|||   :::   :::         :::|||   ::::::
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
              ::::::      ||||||::::::||||||         :::|||:::|||||||||
Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
              ||||||          |||   ||||||::::::|||||||||   ||||||||||||
Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 602
              ||||||||||||||||||||   ||||||:::
Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        602 ------------------------------------------------------------ 602

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        603 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 638
                    |||:::                 :::|||   |||||||||   ||||||
Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
              :::       :::  ||||||||||||||   |||   ||||||            |||
Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
              ||||||   |||||||||   :::   |||   ||||||||||||||            |||
Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
              ||||||   |||   ||||||:::|||   |||:::|||      ::::::   ||||||
Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
              ||||||:::::::::|||||||||||||||   :::::::::||||||||||||
```

```
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
                  :::::::::   ||||||:::   :::|||          |||:::   :::
Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
                :::         |||:::||||||:::|||   |||:::   |||:::|||
Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1057
                |||   ||||||   |||||||||   ||||||   ||||||:::||||||||||||
Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1058 ----------------------------------------------------AAACCCACC 1066
                                                                    :::   |||
Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123
                |||   |||   |||   |||:::     :::|||   :::|||:::   ::::::
Db        396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414

RESULT 11
US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 39:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 415 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

Alignment Scores:
Pred. No.:              1.5e-63        Length:        415
```

```
Score:                  651.50         Matches:       142
Percent Similarity:     52.03%         Conservative:  76
Best Local Similarity:  33.89%         Mismatches:    140
Query Match:            23.46%         Indels:        62
DB:                     4              Gaps:          8

US-09-980-881A-1 (1-1573) x US-09-011-769A-39 (1-415)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 89
              |||   |||||||||   :::   |||            |||            |||:::
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         90 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 149
                 :::|||                        :::   :::::::::
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        150 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 209
              |||   :::|||   :::|||      |||:::|||   :::         |||
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        210 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 269
                    |||   |||   |||   |||   |||         |||:::      |||
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        270 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 329
                 :::   :::          ||||||:::::::::::   :::::::::::
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        330 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
              |||       :::                |||||:::         ||||||:::|||:::
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
                       |||   :::|||   :::   :::         :::|||   ::::::
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
              ::::::      ||||||:::::||||||         :::|||:::|||||||||
Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
              ||||||             |||   ||||||::::::|||||||||   ||||||||||
Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 602
              ||||||||||||||||||||   ||||||:::
Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        602 ------------------------------------------------------------ 602

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        603 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 638
                    |||:::                 ::::|||   |||||||||   ||||||
Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
              :::       :::   ||||||||||||||   |||   ||||||          |||
Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
              ||||||   |||||||||   :::   |||   ||||||||||||||          |||
Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
              ||||||   |||   ||||||:::|||   |||:::|||       ::::::   ||||||
Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
              ||||||:::::::::|||||||||||||||   :::::::::|||||||||||
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
```

```
Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
                  :::         |||:::||||||:::|||   |||:::   |||:::|||
Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1057
                |||   ||||||   |||||||||   ||||||   ||||||:::||||||||||||
Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1058 ----------------------------------------------------AAACCCACC 1066
                                                                    :::   |||
Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123
                |||   |||   |||   |||:::      :::|||   :::|||:::   ::::::
Db        396 CysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHisLeu 414

RESULT 12
US-07-649-591B-6
; Sequence 6, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-6

Alignment Scores:
Pred. No.:              4.17e-63        Length:         417
Score:                  647.50          Matches:        143
Percent Similarity:     52.46%          Conservative:   81
Best Local Similarity:  33.49%          Mismatches:     132
Query Match:            23.32%          Indels:         72
DB:                     1               Gaps:           10

US-09-980-881A-1 (1-1573) x US-07-649-591B-6 (1-417)

Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
                :::   ||||||:::         :::||||||:::      |||:::

Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal---------- 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                              |||         :::|||           |||:::
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
                :::|||      :::::::::::|||       |||   |||:::      |||   |||
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                          :::                |||   |||   |||:::      :::
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                    :::::::::   |||:::   :::   :::         :::|||:::
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
                |||:::::::::   |||::::::|||            :::   :::      |||
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
                |||         |||   :::|||::::::      :::|||   :::|||   |||
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
                :::   :::::::::|||:::|||::::::::::|||   |||||||||:::   |||
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                  |||||||||||||||||:::   |||::::::         :::   |||||||:::
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
                :::|||||||||||||||||||||||||::|||||||||||||||   |||||||:::
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ---------------------------------------------------------- 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
                                               |||:::                 |||
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
                |||||||||||||||||||||||         |||:::   ||||||||||||||||
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
                |||   ||||||   :::       |||                 ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
                :::   |||   |||       ||||||||||   |||   |||||||   :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
                :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
                ::::::|||||||||   ||||||               :::||||||:::|||:::
Db        310 LeuLeuProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
                ||||||                  :::       |||         |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
                |||         |||:::|||         |||         ||||||   ||||||
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
```

```
Qy       1032 TTGGGCATCAAATATTCGTTTAC------------------------------------ 1054
              ||||||||||||:::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                           ||||||||||||||||||||          ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415
```

RESULT 13
US-08-277-540-6
; Sequence 6, Application US/08277540
; Patent No. 5474901
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5474901el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/277,540
; FILING DATE: 19-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-277-540-6

Alignment Scores:

Pred. No.:	4.17e-63	Length:	417
Score:	647.50	Matches:	143
Percent Similarity:	52.46%	Conservative:	81
Best Local Similarity:	33.49%	Mismatches:	132
Query Match:	23.32%	Indels:	72
DB:	1	Gaps:	10

US-09-980-881A-1 (1-1573) x US-08-277-540-6 (1-417)

```
Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
                :::   ||||||:::         :::||||||:::      |||:::

Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                           |||         :::|||           |||:::
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
              :::|||      ::::::::::::|||      |||   |||:::      |||   |||
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                        :::                 |||   |||   |||:::      :::
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                    :::::::::   |||:::   :::   :::         :::|||:::
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
              |||:::::::::   |||::::::|||            :::   :::       |||
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              |||         |||   :::|||::::::      :::|||   :::|||   |||
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              :::   :::::::::|||:::|||:::::::::|||   |||||||||:::   |||
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                  ||||||||||||||||:::   |||::::::         :::   ||||||:::
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
              :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT----------------AAT 611
                                             |||:::                |||
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              |||||||||||||||||||||||          |||:::   ||||||||||||||||
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   ||||||   :::      |||              ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||       ||||||||||   |||   |||||||||   :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::|||||||||   ||||||                :::||||||:::|||:::
Db        310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                :::      |||         |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
              |||            |||:::|||         |||         ||||||   ||||||
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368
```

```
Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
              |||||||||||||:::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                           ||||||||||||||||||||        ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415
```

RESULT 14
US-08-430-787A-6
; Sequence 6, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-6

Alignment Scores:

Pred. No.:	4.17e-63	Length:	417
Score:	647.50	Matches:	143
Percent Similarity:	52.46%	Conservative:	81
Best Local Similarity:	33.49%	Mismatches:	132
Query Match:	23.32%	Indels:	72
DB:	1	Gaps:	10

US-09-980-881A-1 (1-1573) x US-08-430-787A-6 (1-417)

```
Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
              :::  ||||||:::        :::|||||||:::      |||:::
Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                          |||         :::|||           |||:::
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
              :::|||      ::::::::::::|||      |||   |||:::      |||   |||
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                        :::                |||   |||   |||:::      :::
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                  :::::::::   |||:::   :::   :::          :::|||:::
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
              |||:::::::::   |||::::::|||             :::   :::      |||
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              |||         |||   :::|||:::::      :::|||   :::|||   |||
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              :::   :::::::::|||:::|||::::::::::|||   |||||||||:::   |||
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                 ||||||||||||||||||:::   |||::::::         :::   ||||||:::
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
              :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
                                             |||:::                 |||
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              |||||||||||||||||||||||||          |||:::   ||||||||||||||||
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   ||||||   :::      |||                  ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||       ||||||||||||   |||   |||||||||   :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::|||||||||   ||||||                :::|||||:::|||:::
Db        310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                  :::      |||         |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
```

```
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 15
US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 56:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 424 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56

Alignment Scores:
Pred. No.:                1.03e-62        Length:          424
Score:                    644.00          Matches:         136
Percent Similarity:       52.62%          Conservative:    75
Best Local Similarity:    33.92%          Mismatches:      130
Query Match:              23.19%          Indels:          61
DB:                       4               Gaps:            7

US-09-980-881A-1 (1-1573) x US-09-011-769A-56 (1-424)

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
Db         29 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIle 48

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
Db         49 ArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIle 68

Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
Db         69 LysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsn 88

Qy        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
Db         89 ValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValVal 108

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
Db        109 GluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLys 125

Qy        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
Db        126 TyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAla 145

Qy        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
Db        146 LeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLys 165

Qy        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
Db        166 Val---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAla 184

Qy        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
Db        185 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyr 204

Qy        602 ------------------------------------------------------------ 602
Db        205 GlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProVal 224

Qy        603 ------------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGT 632
Db        225 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArg 244

Qy        633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
Db        245 SerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAla 263

Qy        693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
Db        264 GlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAla 283

Qy        753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
Db        284 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSer 303

Qy        813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
Db        304 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyr 323

Qy        873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
Db        324 AlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLys 343

Qy        933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
Db        344 GluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 362

Qy        993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
Db        363 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPhe 382

Qy       1053 ACATC----------------------------------------------------AAA 1060
```

```
              |||                                                  :::
Db         383 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArg 402

Qy        1061 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
               ||||||   |||   |||   |||:::      :::|||   :::|||:::   :::
Db         403 AlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHis 422

Qy        1121 GTT 1123
               :::
Db         423 Leu 423


Search completed: January  1, 2004, 19:03:08
Job time : 69.3933 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:59:01 ; Search time 69.6674 Seconds
(without alignments)
9060.445 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 1467874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135753_25135/app_query.fasta_1.2958
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980881_@CGN_1_1_148_@runat_31122003_135753_25135
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query % Match	Length	DB	ID	Description
1	1885	67.9	423	12	US-10-379-836-17	Sequence 17, Appl
2	1796	64.7	423	12	US-10-379-836-2	Sequence 2, Appli
3	1592.5	57.3	422	12	US-10-379-836-18	Sequence 18, Appl
4	1568.5	56.5	422	12	US-10-379-836-16	Sequence 16, Appl
5	889	32.0	211	9	US-09-925-302-467	Sequence 467, App
6	647.5	23.3	417	12	US-10-341-434-188	Sequence 188, App
7	647.5	23.3	417	15	US-10-229-546-2	Sequence 2, Appli
8	647.5	23.3	417	15	US-10-229-546-9	Sequence 9, Appli
9	644	23.2	402	12	US-10-379-836-20	Sequence 20, Appl
10	635.5	22.9	437	12	US-10-274-639-12	Sequence 12, Appl
11	635.5	22.9	437	14	US-10-200-344-10	Sequence 10, Appl
12	634.5	22.8	716	10	US-09-910-059-125	Sequence 125, App
13	571.5	20.6	613	10	US-09-910-059-113	Sequence 113, App
14	570	20.5	374	9	US-09-888-615-61	Sequence 61, Appl
15	563	20.3	444	16	US-10-176-306-74	Sequence 74, Appl
16	546.5	19.7	399	15	US-10-200-910-8	Sequence 8, Appli
17	529.5	19.1	350	14	US-10-200-344-12	Sequence 12, Appl
18	520	18.7	436	15	US-10-200-910-6	Sequence 6, Appli
19	512	18.4	436	14	US-10-076-535-2	Sequence 2, Appli
20	508	18.3	417	9	US-09-923-779-150	Sequence 150, App
21	504	18.1	428	9	US-09-925-297-528	Sequence 528, App
22	497	17.9	298	12	US-10-379-836-19	Sequence 19, Appl
23	489.5	17.6	406	9	US-09-925-297-517	Sequence 517, App
24	489.5	17.6	419	12	US-10-345-680-65	Sequence 65, Appl
25	483	17.4	421	11	US-09-946-374-234	Sequence 234, App
26	483	17.4	421	12	US-10-015-387A-234	Sequence 234, App
27	483	17.4	421	12	US-10-006-130A-234	Sequence 234, App
28	483	17.4	421	12	US-10-199-672-308	Sequence 308, App
29	483	17.4	421	12	US-10-006-172A-234	Sequence 234, App
30	483	17.4	421	12	US-10-187-749-308	Sequence 308, App
31	483	17.4	421	12	US-10-194-457-308	Sequence 308, App
32	483	17.4	421	12	US-10-184-642-308	Sequence 308, App
33	483	17.4	421	12	US-10-196-747-308	Sequence 308, App
34	483	17.4	421	12	US-10-015-392A-234	Sequence 234, App
35	483	17.4	421	12	US-10-017-253A-234	Sequence 234, App
36	483	17.4	421	12	US-10-173-689-308	Sequence 308, App
37	483	17.4	421	12	US-10-173-690-308	Sequence 308, App
38	483	17.4	421	12	US-10-173-691-308	Sequence 308, App
39	483	17.4	421	12	US-10-173-692-308	Sequence 308, App
40	483	17.4	421	12	US-10-173-694-308	Sequence 308, App
41	483	17.4	421	12	US-10-173-698-308	Sequence 308, App
42	483	17.4	421	12	US-10-173-699-308	Sequence 308, App
43	483	17.4	421	12	US-10-173-707-308	Sequence 308, App
44	483	17.4	421	12	US-10-174-569-308	Sequence 308, App
45	483	17.4	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Alignment Scores:
Pred. No.: 5.6e-179 Length: 423
Score: 1885.00 Matches: 365

```
Percent Similarity:     86.52%          Conservative:  1
Best Local Similarity:  86.29%          Mismatches:    3
Query Match:            67.88%          Indels:        55
DB:                     12              Gaps:          2

US-09-980-881A-1 (1-1573) x US-10-379-836-17 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ------------------------------------------AATCGAATGTGGAGAAAG 626
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Papio hamadryas
US-10-379-836-2

Alignment Scores:
Pred. No.:              4.23e-170       Length:         423
Score:                  1796.00         Matches:        347
Percent Similarity:     84.16%          Conservative:   9
Best Local Similarity:  82.03%          Mismatches:     13
Query Match:            64.67%          Indels:         55
DB:                     12              Gaps:           2

US-09-980-881A-1 (1-1573) x US-10-379-836-2 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||   |||||||||||||||||||||||||||||:::||||||||||||||||||
Db         61 LeuIleGluLysLysLysGlnValHisPhePheValAsnSerSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
```

```
Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              |||||||||||||||||||||||||||||||||||   |||||||||::::::
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluLeuIleThrGluLysTyr 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              |||||||||||||||||||||||||||||||:::|||||:::|||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerTyrGluLysHisProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              |||||||||||||||||||||||||||||||:::|||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaMetTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGlu 200

Qy        608 ------------------------------------------------------------ 608

Db        201 TyrTyrGlyIleIleGlyGluTyrThrAsnLeuLeuArgHisValAspPheTyrValMet 220

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||||||||||||||||   ||||||||||||||||||||   ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnArgCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              |||||||||||||||||||||||   ||||||||||:::|||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluAlaLysAlaValAlaAsnPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||   |||||||||||||   |||||||||||||   |||||||||
Db        341 ValArgAlaIleGlnLysThrSerLysAsnIleArgTyrThrHisGlyArgGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              |||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyAlaAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT-------------------------------------------------------TAC 1054
              |||                                                         |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyLysTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              |||||||||||||||::::::|||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysLysAspAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 3
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-379-836-18

Alignment Scores:
Pred. No.:              8.43e-150       Length:         422
Score:                  1592.50         Matches:        309
Percent Similarity:     79.15%          Conservative:   25
Best Local Similarity:  73.22%          Mismatches:     33
Query Match:            57.35%          Indels:         56
DB:                     12              Gaps:           3

US-09-980-881A-1 (1-1573) x US-10-379-836-18 (1-422)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||||||||       |||   :::||||||   |||:::|||:::   |||||||||
Db          1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db         20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
                 |||   |||||||||:::|||||||||||||||||||||||||||||||:::|||
Db         60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||   ||||||   :::|||||:::   :::|||||||||
Db         80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||:::||||||      ||||||||||||||||||||||||||||||||||||||
Db        100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              |||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db        120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              |||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy        608 ------------------------------------------------------------ 608

Db        200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239
```

```
Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
             |||||||||   :::   ||||||   |||:::|||||||||||||||||   ||||||
Db       240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
             ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db       260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 279

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
             ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
             :::   ||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db       300 AspHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
             ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
             ||||||||||||:::      :::||||||||||||||||||||||||   ||||||||
Db       340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
             :::||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db       360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy      1047 TCGTT----------------------------------------------------TAC 1054
             |||                                                      |||
Db       380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399

Qy      1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
             ||||||||||||||   ||||||   ||||||:::|||||||||   ||||||||||||
Db       400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLysIleValTrpHisValIle 419

Qy      1115 AGGAAT 1120
             ||||||
Db       420 ArgAsn 421
```

RESULT 4
US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-379-836-16

Alignment Scores:

Pred. No.:	2.09e-147	Length:	422
Score:	1568.50	Matches:	305
Percent Similarity:	78.44%	Conservative:	26
Best Local Similarity:	72.27%	Mismatches:	36
Query Match:	56.48%	Indels:	56
DB:	12	Gaps:	3

US-09-980-881A-1 (1-1573) x US-10-379-836-16 (1-422)

```
Qy        18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
             |||||||||      |||   |||||||||   |||:::|||:::   |||:::|||
Db         1 MetLysLeuTyrGlyLeuGlyValLeuValAlaIleIleLeuTyr---GluLysHisGly 19

Qy        78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
             ||||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||
Db        20 LeuAlaPheGlnSerGlyHisValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy       138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
             :::||||||||||||||||||||||||||:::||||||||||||||||||||||||:::
Db        40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy       198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
             |||   |||||||||:::|||||||||||||||||||||||||||||:::::|||
Db        60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAsnSerVal 79

Qy       258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
             ||||||:::||||||   |||   ||||||   :::||||||:::   :::|||||||||
Db        80 LysAlaTyrLeuAsnAlaSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
             |||||||||||||   ||||||||||||||||||||||||||:::|||||||||
Db       100 LeuIleGlnGlnGlnThrSerAsnAspThrValSerProArgAlaSerSerSerTyrTyr 119

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
             ||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db       120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy       438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
             ||||||||||||   ||||||:::|||||||||||:::||||||||||||||||||||||
Db       140 ProAspMetLeuGlnLysIleTyrIleGlySerSerTyrGluLysTyrProLeuTyrVal 159

Qy       498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
             ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db       160 LeuLysValSerGlyLysGluHisArgValLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy       558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy       608 ------------------------------------------------------------ 608

Db       200 PheHisGlyLysGluAsnThrTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy       609 -----------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db       220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
             |||||||||   :::   ||||||   |||:::|||||||||||||||||   ||||||
Db       240 AsnArgSerValHisMetAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
             ||||||||||||||||||:::||||||||||||   ||||||||||||||||||||||||
Db       260 SerLysHisTrpCysGluLysGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 279

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
             ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
             |||   ||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db       300 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
             ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
             ||||||||||||:::      :::||||||||||||||||||||||||   ||||||||
Db       340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
             :::||||||||||||||||||||   ||||||||||||||||||||||||||||||||
```

```
Db        360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy       1047 TCGTTTACATC---------------------------------------------------- 1057
              |||||||||
Db        380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgPhe 399

Qy       1058 ---AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
                 ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||
Db        400 IleLysProThrCysAlaGluAlaLeuAlaAlaValSerLysIleAlaTrpHisValIle 419

Qy       1115 AGGAAT 1120
              ||||||
Db        420 ArgAsn 421

RESULT 5
US-09-925-302-467
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
;   LENGTH: 211
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-302-467

Alignment Scores:
Pred. No.:              1.03e-79       Length:         211
Score:                  889.00         Matches:        173
Percent Similarity:     83.33%         Conservative:   2
Best Local Similarity:  82.38%         Mismatches:     9
Query Match:            32.01%         Indels:         27
DB:                     9              Gaps:           3

US-09-980-881A-1 (1-1573) x US-09-925-302-467 (1-211)

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------------GGC 605
              |||||||||          |||   |||   :::
Db          3 TrpIleSer---MetLeuCysArgTrpLeuMetTrpMetValMetAsnTyrSerTrpLys 21

Qy        606 CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACA 665
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         22 LysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThr 41

Qy        666 GACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCA 725
              |||||||||   ||||||   ||||||||||||||||||||||||||||||||||||||
Db         42 AspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSer 61

Qy        726 TGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCT 785
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAla 81

Qy        786 AGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCC 845
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 SerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSer 101

Qy        846 CAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 GlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 121

Qy        906 TCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTAT 965
              |||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        122 SerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyr 141

Qy        966 ACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATC 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 ThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIle 161

Qy       1026 TATGATTTGGGCATCAAATATTCGTT-------------------------------- 1051
              ||||||||||||||||||||||||||
Db        162 TyrAspLeuGlyIleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGly 181

Qy       1052 ------------------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCT 1093
                                 ||||||||||||||||||||||||||||||||||||||||
Db        182 PheLeuLeuProGluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSer 201

Qy       1094 AAAATAGCTTGGCATGTCATTAGGAATGTT 1123
              |||||||||||||||||||||||||||||
Db        202 LysIleAlaTrpHisValIleArgAsnVal 211

RESULT 6
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-341-434-188

Alignment Scores:
Pred. No.:              1.55e-55       Length:         417
Score:                  647.50         Matches:        143
Percent Similarity:     52.46%         Conservative:   81
Best Local Similarity:  33.49%         Mismatches:     132
Query Match:            23.32%         Indels:         72
DB:                     12             Gaps:           10

US-09-980-881A-1 (1-1573) x US-10-341-434-188 (1-417)

Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
              :::   ||||||:::         :::||||||:::      |||:::
Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                                 |||     :::|||          |||:::
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
              :::|||      ::::::::::::|||       |||   |||:::      |||  |||
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                          :::                |||   |||   |||:::      :::
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                    :::::::::   |||:::   :::   :::        :::|||:::
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
              |||:::::::::   |||::::::|||            :::   :::      |||
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114
```

```
Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
Db        310 LeuLeuProProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 7
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-2

Alignment Scores:
Pred. No.:                1.55e-55        Length:         417
Score:                    647.50          Matches:        143
Percent Similarity:       52.46%          Conservative:   81
Best Local Similarity:    33.49%          Mismatches:     132
Query Match:              23.32%          Indels:         72
DB:                       15              Gaps:           10

US-09-980-881A-1 (1-1573) x US-10-229-546-2 (1-417)

Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
```

```
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   ||||||  :::      |||                 ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||      ||||||||||   |||   ||||||||  :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::||:::||||||   ||||||:::   ||||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::|||||||||   ||||||                ::::||||||:::|||:::
Db        310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                  :::      |||         |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
              |||           |||:::|||         |||         ||||||  ||||||
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
              ||||||||||||:::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 ---------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                            ||||||||||||||||||||         ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 8
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-9

Alignment Scores:
Pred. No.:              1.55e-55        Length:         417
Score:                  647.50          Matches:        143
Percent Similarity:     52.46%          Conservative:   81
Best Local Similarity:  33.49%          Mismatches:     132
Query Match:            23.32%          Indels:         72
DB:                     15              Gaps:           10

US-09-980-881A-1 (1-1573) x US-10-229-546-9 (1-417)

Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65

Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal---------- 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                        |||      ::|||          |||:::
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
              :::|||      ::::::::::::|||      |||   |||:::      |||   |||
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                          :::                 |||   |||   |||:::      :::
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                    :::::::::   |||:::   :::   :::       :::|||:::
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
              |||:::::::::   |||::::::|||           :::   :::     |||
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              |||          |||   :::|||::::::      :::|||   :::|||   |||
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              :::   :::::::::|||:::|||:::::::::|||   ||||||||:::   |||
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                 ||||||||||||||||||:::   |||::::::          :::   ||||||:::
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
              :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
                                             |||:::                 |||
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              ||||||||||||||||||||||||      |||:::   ||||||||||||||||||
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   ||||||  :::      |||                 ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||      ||||||||||   |||   ||||||||  :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::|||||||||   ||||||                ::::||||||:::|||:::
Db        310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                  :::      |||         |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
              |||           |||:::|||         |||         ||||||  ||||||
```

```
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
              |||||||||||||:::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                           ||||||||||||||||||||          ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 9
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-379-836-20

Alignment Scores:
Pred. No.:                3.43e-55        Length:         402
Score:                    644.00          Matches:        136
Percent Similarity:       52.62%          Conservative:   75
Best Local Similarity:    33.92%          Mismatches:     130
Query Match:              23.19%          Indels:         61
DB:                       12              Gaps:           7

US-09-980-881A-1 (1-1573) x US-10-379-836-20 (1-402)

Qy          84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
               |||:::     :::|||                                 :::   ::::::
Db           7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIle 26

Qy         144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
               :::   |||   :::|||   ::|||       |||:::|||   :::          |||
Db          27 ArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIle 46

Qy         204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
                         |||   |||   |||   |||   |||          |||:::
Db          47 LysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsn 66

Qy         264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
                  |||       :::   :::          ||||||:::::::::::    :::::::::
Db          67 ValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValVal 86

Qy         324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
               :::   |||       :::          ||||||:::       ||||||:::
Db          87 GluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLys 103

Qy         384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
               |||:::             |||   :::|||   :::   :::          :::|||
Db         104 TyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAla 123

Qy         444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
               ::::::::::::      ||||||::::::||||||         :::|||:::||||||
Db         124 LeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLys 143

Qy         504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
               |||   ||||||            |||   ||||||::::::|||||||||   ||||||
Db         144 Val---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162

Qy         564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
               ||||||||||||||||||||||||||   ||||||:::
Db         163 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyr 182

Qy         602 ------------------------------------------------------------ 602

Db         183 GlyArgGluIleGlnValThrGluLeuLeuAsnLysLeuAspPheTyrValLeuProVal 202

Qy         603 ------------GGCCAT------------------AATCGAATGTGGAGAAAGAACCGT 632
                         |||:::                  :::|||   |||||||||   |||
Db         203 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArg 222

Qy         633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
               |||   :::       :::   ||||||||||||||   |||   ||||||
Db         223 SerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAla 241

Qy         693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
               |||||||||   |||||||||   :::   |||   ||||||||||||||
Db         242 GlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAla 261

Qy         753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
               |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db         262 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSer 281

Qy         813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
               ||||||||||||:::::::::|||||||||||||||   :::::::::||||||||||
Db         282 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyr 301

Qy         873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
                         :::::::::   ||||||:::   :::|||         |||:::
Db         302 AlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLys 321

Qy         933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
               :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db         322 GluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 340

Qy         993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
               |||   |||   ||||||   |||||||||   ||||||   ||||||:::|||||||||
Db         341 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPhe 360

Qy        1053 ACATC----------------------------------------------------AAA 1060
               |||                                                      :::
Db         361 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArg 380

Qy        1061 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
               ||||||   |||   |||   |||:::      :::|||   :::|||:::   :::
Db         381 AlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGluHis 400

Qy        1121 GTT 1123
               :::
Db         401 Leu 401

RESULT 10
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
```

```
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 12
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

Alignment Scores:
Pred. No.:              2.48e-54        Length:         437
Score:                  635.50          Matches:        147
Percent Similarity:     49.66%          Conservative:   70
Best Local Similarity:  33.64%          Mismatches:     147
Query Match:            22.88%          Indels:         74
DB:                     12              Gaps:           12

US-09-980-881A-1 (1-1573) x US-10-274-639-12 (1-437)

Qy          9 GCTGTTGGGATGAAGCTTTGC------AGCCTTGCAGTCCTTGTACCCATTGTTCTCTTC 62
              |||      :::   ||||||         |||   :::|||   |||
Db         12 AlaAlaPheLeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro------------ 27

Qy         63 TGTGAGCAGCATGTCTTC------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCT 116
                 |||::::::         :::         :::|||:::      :::|||
Db         28 GlyHisSerHisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIlePro 47

Qy        117 AGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC 176
              :::|||      :::         |||:::   ::::::         ::::::   |||
Db         48 LysThrGluGluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeu 67

Qy        177 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAG------AAAAAACAAGTCCATTTTTTT 230
              |||||||||   :::         :::   :::                 ||||||
Db         68 TrpGlnProSerSerIleSerTyrValSerGluGlyThrValThrAspValHisIlePro 87

Qy        231 GTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 290
                |||   |||             :::   |||   |||      :::   |||
Db         88 GlnAsnGlySerArgAla------LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyr 105

Qy        291 AGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 350
              ||||||:::   |||::::::      :::::::::      |||      |||
Db        106 LysValLeuIleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGln 125

Qy        351 AGCCCCCGAGCCTCCGCATCG------TACTATGAACAGTATCACTCACTAAATGAAATC 404
                  |||   |||   |||           ||||||   ||||||||||||   ||||||
Db        126 ArgAsnArgArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIle 145

Qy        405 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 464
                 :::|||:::      :::   :::   |||      ::::::           |||
Db        146 GlnAsnTrpMetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIle 165

Qy        465 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA 524
              |||   |||:::|||         |||::::::||||||:::   |||:::
Db        166 GlyArgSerTyrGluGlyArgSerLeuPheIleLeuLysLeu---GlyArgArgSerArg 184

Qy        525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
                 |||   |||:::||||||||||||||||||||||||||||||||||||   ||||||
Db        185 LeuLysArgAlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAla 204

Qy        585 TTCTGCTTGTGGTTCATA---------------------------------------- 602
              ||||||   ||||||:::
Db        205 PheCysGlnTrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArg 224

Qy        602 ---------------------------------------------------------- 602

Db        225 LysMetLeuAsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHis 244

Qy        603 ---------GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 653
                      ::::::|||   |||||||||   ||||||      :::
Db        245 PheSerTrpThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArg 264

Qy        654 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 713
              |||   |||   |||   |||   |||:::      |||   ||||||:::|||||||||
Db        265 CysArgGlyValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAla 283

Qy        714 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 773
              |||         |||   :::||||||||||||   :::||||||||||||||||||||
Db        284 SerMetHisProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluVal 303

Qy        774 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 833
              ||||||||||||:::|||||||||:::::          |||:::||||||:::|||
Db        304 LysAlaValAlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPhe 323

Qy        834 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 893
              |||:::|||:::|||   ::::::::|||||||||||      :::         :::
Db        324 HisAlaTyrAlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsn 343

Qy        894 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 953
                    :::         |||   :::||||||   |||:::   ::::::
Db        344 PheArgCysValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyr 362

Qy        954 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 1013
                    ||||||   :::|||      |||   |||||||||::::::   |||
Db        363 GlyValArgTyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSer 382

Qy       1014 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC------------------ 1054
              ||||||   |||       ||||||   |||:::|||
Db        383 MetAspTrpAlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThr 402

Qy       1055 --------------------------------ATCAAACCCACCTGTAGAGAAGCTTTT 1081
                                               ||||||||||||||   |||
Db        403 GlyTyrPheGlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMet 422

Qy       1082 GCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCC 1132
                 ||||||      |||      |||:::::::::          ||||||
Db        423 LeuAlaValLysAsnIleThrMetHisLeuLeuLysLys------CysPro 437

RESULT 11
US-10-200-344-10
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
```

```
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-200-344-10

Alignment Scores:
Pred. No.:              2.48e-54      Length:          437
Score:                  635.50        Matches:         147
Percent Similarity:     49.66%        Conservative:    70
Best Local Similarity:  33.64%        Mismatches:      147
Query Match:            22.88%        Indels:          74
DB:                     14            Gaps:            12

US-09-980-881A-1 (1-1573) x US-10-200-344-10 (1-437)

Qy          9 GCTGTTGGGATGAAGCTTTGC------AGCCTTGCAGTCCTTGTACCCATTGTTCTCTTC 62
              |||      :::   ||||||         |||   :::|||   |||
Db         12 AlaAlaPheLeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro------------ 27

Qy         63 TGTGAGCAGCATGTCTTC------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCT 116
                 |||::::::         :::         :::|||:::      :::|||
Db         28 GlyHisSerHisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIlePro 47

Qy        117 AGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC 176
              :::|||      :::         |||:::   ::::::        ::::::   |||
Db         48 LysThrGluGluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeu 67

Qy        177 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAG------AAAAAACAAGTCCATTTTTTT 230
              |||||||||   :::         :::   :::             ||||||
Db         68 TrpGlnProSerSerIleSerTyrValSerGluGlyThrValThrAspValHisIlePro 87

Qy        231 GTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 290
                 |||   |||           :::   |||   |||      :::   |||
Db         88 GlnAsnGlySerArgAla------LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyr 105

Qy        291 AGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 350
              |||||:::   |||::::::      :::::::::      |||        |||
Db        106 LysValLeuIleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGln 125

Qy        351 AGCCCCCGAGCCTCCGCATCG------TACTATGAACAGTATCACTCACTAAATGAAATC 404
                |||   |||   |||         ||||||   ||||||||||||   ||||||
Db        126 ArgAsnArgArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIle 145

Qy        405 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 464
                :::|||:::      :::   :::   |||       ::::::             |||
Db        146 GlnAsnTrpMetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIle 165

Qy        465 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA 524
              |||   |||:::|||         |||::::::||||||:::   |||:::
Db        166 GlyArgSerTyrGluGlyArgSerLeuPheIleLeuLysLeu---GlyArgArgSerArg 184

Qy        525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
                |||   |||:::||||||||||||||||||||||||||||||||||   ||||||
Db        185 LeuLysArgAlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAla 204

Qy        585 TTCTGCTTGTGGTTCATA---------------------------------------- 602
              ||||||   ||||||:::
Db        205 PheCysGlnTrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArg 224

Qy        602 ------------------------------------------------------------ 602

Db        225 LysMetLeuAsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHis 244

Qy        603 ---------GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 653
                      ::::::|||   |||||||||   ||||||      :::

Db        245 PheSerTrpThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArg 264

Qy        654 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 713
              |||   |||   |||   |||   |||:::      |||   ||||||:::|||||||||
Db        265 CysArgGlyValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAla 283

Qy        714 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 773
              |||         |||   :::||||||||||||   :::||||||||||||||||||||
Db        284 SerMetHisProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluVal 303

Qy        774 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 833
              ||||||||||||:::|||||||||::::::         |||:::||||||:::|||
Db        304 LysAlaValAlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPhe 323

Qy        834 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 893
              |||:::|||:::|||   :::::::::||||||||||||      :::         :::
Db        324 HisAlaTyrAlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsn 343

Qy        894 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 953
                      :::         |||   :::||||||   |||:::   ::::::
Db        344 PheArgCysValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyr 362

Qy        954 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 1013
                  ||||||   :::|||       |||   ||||||||:::::   |||
Db        363 GlyValArgTyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSer 382

Qy       1014 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC------------------- 1054
                ||||||   |||       ||||||   |||:::|||
Db        383 MetAspTrpAlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThr 402

Qy       1055 --------------------------------ATCAAACCCACCTGTAGAGAAGCTTTT 1081
                                               ||||||||||||||   |||
Db        403 GlyTyrPheGlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMet 422

Qy       1082 GCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCC 1132
                ||||||      |||      |||:::::::::          ||||||
Db        423 LeuAlaValLysAsnIleThrMetHisLeuLeuLysLys------CysPro 437

RESULT 12
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody, a
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-125

Alignment Scores:
Pred. No.:              3.61e-54      Length:          716
Score:                  634.50        Matches:         139
Percent Similarity:     51.66%        Conservative:    79
```

```
Best Local Similarity:  32.94%        Mismatches:    143
Query Match:            22.85%        Indels:         62
DB:                     10            Gaps:            8

US-09-980-881A-1 (1-1573) x US-09-910-059-125 (1-716)

Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
              :::   :::|||   |||||||||   :::   |||           |||
Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         84 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
                 |||:::      :::|||                              :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy        141 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 200
              ::::::   |||   :::|||   :::|||      |||:::|||   :::
Db         41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        201 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 260
              |||                |||   |||   |||   |||   |||           |||:::
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy        261 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 320
                 |||      :::   :::         ||||||:::::::::::::   ::::::
Db         81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy        321 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 380
              ::::::   |||      :::                ||||||:::          ||||||
Db        101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
              :::|||:::              |||   :::|||   :::   :::          :::|||
Db        118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
                 ::::::::::::      ||||||:::::||||||          :::|||:::|||
Db        138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
              ||||||   ||||||            |||   ||||||::::::|||||||||   |||
Db        158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------------ 602
              ||||||||||||||||||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy        602 ---------------------------------------------------------- 602

Db        197 TyrGlyArgGluIleGlnValThrGluLeuAspLysLeuAspPheTyrValLeuPro 216

Qy        603 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629
                            |||:::               :::|||   |||||||||
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy        630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
              ||||||   :::      :::   ||||||||||||||   |||   ||||||
Db        237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
                 |||||||||   |||||||||   :::   |||   |||||||||||||||
Db        256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
                 |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
              |||||||||||:::::::::|||||||||||||||||   :::::::::|||||||||
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
              |||                 :::::::::   ||||||:::   :::|||          |||
Db        316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
                 :::   :::      :::         |||:::||||||:::|||   |||:::   |||
Db        336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
                 :::|||   :::   |||          ||||||   ||||||   ||||||:::||||||
Db        355 IleTyrProSerAlaGlyThrSerLysAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

Qy       1050 TTTACATC-------------------------------------------------- 1057
              ||||||
Db        375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

Qy       1058 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGG 1117
                 :::   ||||||   |||   |||   |||:::      :::|||   :::|||:::
Db        395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 414

Qy       1118 AATGTT 1123
              ::::::
Db        415 HisLeu 416

RESULT 13
US-09-910-059-113
; Sequence 113, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
;   LENGTH: 613
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised Fd mutant HCPB sequence
US-09-910-059-113

Alignment Scores:
Pred. No.:              6.64e-48        Length:         613
Score:                  571.50          Matches:        136
Percent Similarity:     45.58%          Conservative:   60
Best Local Similarity:  31.63%          Mismatches:     118
Query Match:            20.58%          Indels:         117
DB:                     10              Gaps:           11

US-09-980-881A-1 (1-1573) x US-09-910-059-113 (1-613)

Qy         20 GAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTT 79
              |||         |||           ||||||                        |||
Db        238 GluLeuLysThrProLeuGlyAspThrThrHis-----------------ThrCysPro 251

Qy         80 CGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGT 139
              |||   ||||||           ||||||
Db        252 ArgCysProGlu------ProLysSerCys---------------------------- 259

Qy        140 TCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCT 199
                                                                ::::::   |||
Db        260 ---------------------------------------------AspThrProPro 263
```

```
Qy       200 TATTGTGAAGAAAAAACAAGTCCATTTTTTTGT--------------------AAATGC 238
Db       264 ProCysProArgCysProGluProLysSerCysAspThrProProProCysProArgCys 283

Qy       239 ATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTT 298
Db       284 ProGluProLysSerCysAspThrProProProCysProArg-CysProAlaProGluLe 303

Qy       299 GCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCG 358
Db       303 uLeuGlyGly---------------------------------------------------- 306

Qy       359 AGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGA 418
Db       307 -AlaThrGlyHisSerTyrGluLysTyrAsnLysTrpGluThrIleGluAlaTrpThrGl 326

Qy       419 ATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGA 478
Db       326 nGlnValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlyThrThrPheGl 346

Qy       479 GAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCAT 538
Db       346 uGlyArgAlaIleTyrLeuLeuLysVal---GlyLysAlaGlyGlnAsnLysProAlaIl 365

Qy       539 ATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTT 598
Db       365 ePheMetAspCysGlyPheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPh 385

Qy       599 CATA---------------------------------------------------------- 602
Db       385 eValArgGluAlaValArgThrTyrGlyArgGluIleGlnValThrGluLeuLeuAspLy 405

Qy       603 -------------------------------------GGCCAT----------------- 608
Db       405 sLeuAspPheTyrValLeuProValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLy 425

Qy       609 -AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGA 667
Db       425 sSerArgPheTrpArgLysThrArgSerThrHisThrGlySerSerCysIleGlyThrAs 445

Qy       668 CCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATG 727
Db       445 pProAsnArgAsnPhe---AspAlaGlyTrpCysGluIleGlyAlaSerArgAsnProCy 464

Qy       728 CTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAG 787
Db       464 sAspGluThrTyrCysGlyProAlaAlaGluSerGluLysGluThrLysAlaLeuAlaAs 484

Qy       788 TTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCA 847
Db       484 pPheIleArgAsnLysLeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGl 504

Qy       848 GCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTC 907
Db       504 nMetMetIleTyrProTyrSerTyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAs 524

Qy       908 TCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATAC 967
Db       524 nAlaLeuAlaLysAlaThrValLysGluLeu---AlaSerLeuHisGlyThrLysTyrTh 543

Qy       968 ACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTA 1027
Db       543 rTyrGlyProGlyAlaThrThrIleTyrProSerAlaGlyThrSerLysAspTrpAlaTy 563

Qy      1028 TGATTTGGGCATCAAATATTCGTTTACATC----------------------------- 1057
Db       563 rAspGlnGlyIleArgTyrSerPheThrPheGluLeuArgAspThrGlyArgTyrGlyPh 583

Qy      1058 ---------------------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAA 1095
Db       583 eLeuLeuProGluSerGlnIleArgAlaThrCysGluGluThrPheLeuAlaIleLysTy 603

Qy      1096 AATAGCTTGGCATGTCATTAGGAATGTT 1123
Db       603 rValAlaSerTyrValLeuGluHisLeu 612
```

RESULT 14
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY
; APPLICANT: WHYTE, DAVID
; APPLICANT: CAENEPEEL, SEAN
; APPLICANT: CHARYDCZAK, GLEN
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1214
; CURRENT APPLICATION NUMBER: US/09/888,615
; CURRENT FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: 60/214,047
; PRIOR FILING DATE: 2000-06-26
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-61

Alignment Scores:

Pred. No.:	8.09e-48	Length:	374
Score:	570.00	Matches:	117
Percent Similarity:	52.47%	Conservative:	53
Best Local Similarity:	36.11%	Mismatches:	90
Query Match:	20.53%	Indels:	65
DB:	9	Gaps:	7

US-09-980-881A-1 (1-1573) x US-09-888-615-61 (1-374)

```
Qy       324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAA 380
Db        29 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 48

Qy       381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
Db        49 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 68

Qy       441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
Db        69 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu 88

Qy       501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
Db        89 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 108

Qy       561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 602
Db       109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128

Qy       602 ------------------------------------------------------------ 602
Db       129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 148

Qy       603 --------------GGCCAT------------------AATCGAATGTGGAGAAAGAAC 629
Db       149 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 168

Qy       630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
Db       169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187

Qy       690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
```

```
Db        188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
              |||   ||||||||||||   ||||||||||||||||||:::            :::
Db        208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
              |||      :::::::::|||||||||||   |||   |||:::   ||||||
Db        228 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
              ||||||::::::||||||   :::|||   |||:::     |||     :::|||
Db        248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
              |||:::   |||             |||   |||     |||     ::::::
Db        268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 286

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
              ||||||   :::   |||         ||||||      |||:::||||||   :::|||
Db        287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 306

Qy       1050 TTTACATC---------------------------------------------------- 1057
              :::|||
Db        307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326

Qy       1058 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCT--------------- 1102
              :::||||||||||   |||         ||||||     :::
Db        327 GlnProThrCysGluGluThrMetGluAlaValLeuSerValLeuAspAspValTyrAla 346

Qy       1103 ------TGGCAT 1108
                    ||||||
Db        347 LysHisTrpHis 350

RESULT 15
US-10-176-306-74
; Sequence 74, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 444
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-306-74

Alignment Scores:
Pred. No.:                4.25e-47        Length:          444
Score:                    563.00          Matches:         109
Percent Similarity:       56.74%          Conservative:    51
Best Local Similarity:    38.65%          Mismatches:      82
Query Match:              20.27%          Indels:          40
DB:                       16              Gaps:            5

US-09-980-881A-1 (1-1573) x US-10-176-306-74 (1-444)

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAA 380
              :::|||:::|||   :::   :::|||||||||   :::      :::|||   |||
Db        159 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 178

Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
              ||||||   :::   |||||||||   |||:::      |||:::|||::::::
Db        179 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 198

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
              :::::::::|||:::      :::|||   ::::::|||   :::|||:::|||   |||
Db        199 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProIleTyrTyrLeu 218

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
              |||:::|||.           |||      ||||||:::|||||||||||||||||
Db        219 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 238

Qy        561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 602
              |||||||||||||:::|||||||||||   ||||||:::
Db        239 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 258

Qy        602 ---------------------------------------------------------- 602

Db        259 HisLysAspAsnSerArgIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 278

Qy        603 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629
                            |||:::                 :::|||:::|||||||||:::
Db        279 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 298

Qy        630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
              ||||||   :::   |||       |||   ||||||||||||||   ||||||   :::
Db        299 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 317

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
                  ||||||      |||||||||   :::          :::|||:::||||||
Db        318 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 337

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
              |||   ||||||||||||   ||||||||||||||||||:::            :::
Db        338 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 357

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
              |||      :::::::::|||||||||||   |||   |||:::   ||||||
Db        358 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 377

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
              ||||||::::::||||||   :::|||   |||:::     |||     :::|||
```

```
Db        378 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 397

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
                 |||:::   |||           |||   |||      |||      ::::::
Db        398 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 416

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
              ||||||   :::   |||         ||||||      |||:::||||||   :::|||
Db        417 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 436

Qy       1050 TTTACA 1055
              :::|||
Db        437 TyrThr 438
```

Search completed: January 1, 2004, 19:25:30
Job time : 98.6674 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:25:26 ; Search time 30.8315 Seconds
(without alignments)
9812.904 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135749_24966/app_query.fasta_1.2958
-DB=PIR_76 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_71_@runat_31122003_135749_24966 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1885	67.9	423	2	A41204	carboxypeptidase B
2	695.5	25.0	415	2	A32129	carboxypeptidase B
3	655	23.6	417	1	A34487	carboxypeptidase A
4	647.5	23.3	417	2	A43929	carboxypeptidase A
5	634	22.8	416	1	A42332	carboxypeptidase B
6	589.5	21.2	306	1	CPBOB	carboxypeptidase B
7	567.5	20.4	309	2	A38395	mast cell carboxyp
8	553	19.9	417	2	A32128	carboxypeptidase A
9	536	19.3	419	1	CPBOA	carboxypeptidase A
10	529	19.0	419	1	CPRTA	carboxypeptidase A
11	508	18.3	417	2	A56171	carboxypeptidase A
12	489.5	17.6	419	1	S29127	carboxypeptidase A
13	437.5	15.8	528	2	T33527	hypothetical prote
14	400.5	14.4	667	2	T33526	hypothetical prote
15	397	14.3	323	2	T20507	hypothetical prote
16	391.5	14.1	527	2	T27572	hypothetical prote
17	369.5	13.3	303	1	CPCYB	carboxypeptidase B
18	362.5	13.1	455	2	T33572	hypothetical prote
19	358	12.9	666	2	T24170	hypothetical prote
20	299.5	10.8	430	2	S48976	ECM14 protein - ye
21	277.5	10.0	424	2	S17571	carboxypeptidase T
22	277	10.0	451	2	S20723	carboxypeptidase (
23	255.5	9.2	497	2	T40260	carboxypeptidase p
24	244.5	8.8	373	2	T26030	hypothetical prote
25	181.5	6.5	999	2	T36021	probable zinc-bind
26	135.5	4.9	1446	2	T30916	carboxypeptidase D
27	130.5	4.7	558	2	S75104	hypothetical prote
28	118.5	4.3	707	2	T40070	origin recognition
29	118	4.2	1216	2	H84629	hypothetical prote
30	112	4.0	1968	1	PN0093	genome polyprotein
31	105.5	3.8	985	2	D82776	pyruvate dehydroge
32	101.5	3.7	985	2	T29910	hypothetical prote
33	100.5	3.6	537	2	C64432	hypothetical prote
34	100	3.6	376	1	E69957	gamma-D-glutamyl-L
35	99.5	3.6	1389	2	I50090	carboxypeptidase g
36	99	3.6	1587	2	AB2012	hypothetical prote
37	98.5	3.5	606	2	E90536	hypothetical prote
38	98.5	3.5	983	2	B49284	immediate-early pr
39	98.5	3.5	1078	2	T44232	hypothetical prote
40	98.5	3.5	1160	2	I40589	parasporal crystal
41	98	3.5	491	2	B96739	hypothetical prote
42	97.5	3.5	355	2	C83850	gamma-D-glutamyl-L
43	97.5	3.5	7829	2	T15789	hypothetical prote
44	97	3.5	663	2	T37772	telomere length re
45	96	3.5	368	2	T21748	hypothetical prote

ALIGNMENTS

RESULT 1
A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N;Alternate names: plasma carboxypeptidase B
C;Species: Homo sapiens (man)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C;Accession: A41204
R;Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A;Title: Isolation, molecular cloning, and partial characterization of a novel carboxypep
A;Reference number: A41204; MUID:92042093; PMID:1939207
A;Accession: A41204
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <EAT>
A;Cross-references: GB:M75106; NID:g189686; PID:g189687
C;Genetics:
A;Gene: GDB:CPB2
A;Cross-references: GDB:129546; OMIM:212070
A;Map position: 13q14.11-13q14.11
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-114/Domain: activation peptide #status predicted <ACP>
F;115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Alignment Scores:

Pred. No.:	1.36e-158	Length:	423
Score:	1885.00	Matches:	365
Percent Similarity:	86.52%	Conservative:	1
Best Local Similarity:	86.29%	Mismatches:	3
Query Match:	67.88%	Indels:	55
DB:	2	Gaps:	2

US-09-980-881A-1 (1-1573) x A41204 (1-423)

Qy 18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77

```
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 ------------------------------------------AATCGAATGTGGAGAAAG 626
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C;Accession: A32129; S17543
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparat
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: A32129
A;Molecule type: DNA
A;Residues: 1-415 <CLA>
A;Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R;Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A;Title: Distribution of manganese in rat pancreas and identification of its primary binc
A;Reference number: S17543; MUID:91378950; PMID:1898371
A;Accession: S17543
A;Molecule type: protein
A;Residues: 109-116,'X',118-130 <KOD>
C;Genetics:
A;Gene: CPB
A;Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;354,376/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:                2.7e-53        Length:          415
Score:                    695.50         Matches:         158
Percent Similarity:       52.74%         Conservative:    63
Best Local Similarity:    37.71%         Mismatches:      135
Query Match:              25.05%         Indels:          64
DB:                       2              Gaps:            8

US-09-980-881A-1 (1-1573) x A32129 (1-415)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||   |||           |||   |||   :::           |||:::|||
Db          1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGluGluHis--- 19

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
                    |||         :::|||                                |||
Db         20 ------PheAspGlyAsnArgValTyrArgValSerValHisGlyGluAspHisValAsn 37

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ::::::|||   |||       |||   ||||||       |||:::|||   :::|||
Db         38 LeuIleGlnGluLeuAlaAsnThrLysGluIleAspPheTrpLysProAspSerAlaThr 57

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
                 :::          |||   |||   |||   |||   ||||||   :::|||
Db         58 GlnValLysProLeuThrThrValAspPheHisValLysAlaGluAspValAlaAspVal 77

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              :::       |||       :::   :::          ||||||::::::::|||   :::
Db         78 GluAsnPheLeuGluGluAsnGluValHisTyrGluValLeuIleSerAsnValArgAsn 97

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
                 ::::::   |||       :::   |||          |||||||||          |||
```

```
Db         98 AlaLeuGluSerGlnPheAspSerHisThr---------ArgAlaSerGlyHisSerTyr 114

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
Db        115 ThrLysTyrAsnLysTrpGluThrIleGluAlaTrpIleGlnGlnValAlaThrAspAsn 134

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
Db        135 ProAspLeuValThrGlnSerValIleGlyThrThrPheGluGlyArgAsnMetTyrVal 154

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
Db        155 LeuLysIle---GlyLysThrArgProAsnLysProAlaIlePheIleAspCysGlyPhe 173

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------- 602
Db        174 HisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArg 193

Qy        602 ------------------------------------------------------------ 602
Db        194 ThrTyrAsnGlnGluIleHisMetLysGlnLeuLeuAspGluLeuAspPheTyrValLeu 213

Qy        603 ------------------GGCCAT------------------AATCGAATGTGGAGAAAG 626
Db        214 ProValValAsnIleAspGlyTyrValTyrThrTrpThrLysAspArgMetTrpArgLys 233

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
Db        234 ThrArgSerThrMetAlaGlySerSerCysLeuGlyValArgProAsnArgAsnPhe--- 252

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
Db        253 AsnAlaGlyTrpCysGluValGlyAlaSerArgSerProCysSerGluThrTyrCysGly 272

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
Db        273 ProAlaProGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeu 292

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
Db        293 SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrProTyr 312

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db        313 SerTyrAspTyrLysLeuProGluAsnTyrGluGluLeuAsnAlaLeuValLysGlyAla 332

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
Db        333 AlaLysGluLeu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThr 351

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
Db        352 ThrIleTyrProAlaAlaGlyGlySerAspAspTrpSerTyrAspGlnGlyIleLysTyr 371

Qy       1047 TCGTTTACATC---------------------------------------------------- 1057
Db        372 SerPheThrPheGluLeuArgAspThrGlyPhePheGlyPheLeuLeuProGluSerGln 391

Qy       1058 ---AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTC 1111
Db        392 IleArgGlnThrCysGluGluThrMetLeuAlaValLysTyrIleAlaAsnTyrVal 410

RESULT 3
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34487
R;Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A;Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A;Reference number: A34487; MUID:90062123; PMID:2584208
A;Accession: A34487
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:                1.04e-49        Length:          417
Score:                    655.00          Matches:         150
Percent Similarity:       51.67%          Conservative:    66
Best Local Similarity:    35.89%          Mismatches:      138
Query Match:              23.59%          Indels:          65
DB:                       1               Gaps:            9

US-09-980-881A-1 (1-1573) x A34487 (1-417)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 92
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         93 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 152
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        153 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 212
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        213 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 272
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        273 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 332
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        333 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 608
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        609 AAT---------------------------------------------------------- 611
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        612 ------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 638
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278
```

```
Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1054
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1055 ----------------------------------------------ATCAAACCCACC 1066
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
Db        398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415
```

RESULT 4
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N;Alternate names: mast cell carboxypeptidase A3
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: A43929; A39246; A45759
R;Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A;Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A
A;Reference number: A43929; MUID:92105393; PMID:1729276
A;Accession: A43929
A;Molecule type: DNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:M73716
A;Experimental source: mast cell
A;Note: the authors translated the codon CGC for residue 231 as Thr
R;Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafin,
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A;Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison
A;Reference number: A39246; MUID:90083291; PMID:2594780
A;Accession: A39246
A;Molecule type: mRNA
A;Residues: 1-417 <RE2>
A;Cross-references: GB:M27717; NID:g179933; PIDN:AAA35652.1; PID:g179934
R;Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.
J. Clin. Invest. 83, 1630-1636, 1989
A;Title: Human mast cell carboxypeptidase. Purification and characterization.
A;Reference number: A45759; MUID:89214692; PMID:2708524
A;Accession: A45759
A;Molecule type: protein
A;Residues: 110-137 <GOL>
C;Genetics:
A;Gene: GDB:CPA3
A;Cross-references: GDB:125231; OMIM:114851
A;Map position: 3q21.3-3q25
A;Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-109/Domain: activation peptide #status predicted <ACT>
F;110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F;176,179,304/Binding site: zinc (His, Glu, His) #status predicted
F;245-268/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	4.8e-49	Length:	417
Score:	647.50	Matches:	143
Percent Similarity:	52.46%	Conservative:	81
Best Local Similarity:	33.49%	Mismatches:	132
Query Match:	23.32%	Indels:	72
DB:	2	Gaps:	10

US-09-980-881A-1 (1-1573) x A43929 (1-417)

```
Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602
Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 --------------------------------GGCCAT-----------------AAT 611
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
```

```
           :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
           ::::::|||||||||   ||||||              :::||||||:::|||:::
Db        310 LeuLeuPheProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
           ||||||         :::      |||        |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
           |||            |||:::|||         |||         ||||||   ||||||
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
           ||||||||||||:::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 --------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                           ||||||||||||||||||||         ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
           |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 5
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N;Alternate names: carboxypeptidase B isozyme 1
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: A42332; S08254; S02813; S02812
R;Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992
A;Title: Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Ident
A;Reference number: A42332; MUID:92129345; PMID:1370825
A;Accession: A42332
A;Molecule type: mRNA
A;Residues: 1-416 <YAM>
A;Cross-references: GB:M81057; NID:g809194; PIDN:AAA66973.1; PID:g189625
A;Experimental source: pancreas
A;Note: sequence extracted from NCBI backbone (NCBIN:78573, NCBIP:78574)
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08254
A;Molecule type: protein
A;Residues: 16-40 <MOU>
A;Experimental source: pancreas
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02813
A;Molecule type: protein
A;Residues: 16-36,'X',38-43 <PAS>
A;Experimental source: pancreas
A;Accession: S02812
A;Molecule type: protein
A;Residues: 16-36,'Q',38-43 <PA2>
A;Experimental source: pancreas
C;Genetics:
A;Gene: GDB:CPB1
A;Cross-references: GDB:132160; OMIM:114852
A;Map position: 3q21.3-3q25
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-110/Domain: activation peptide #status predicted <ACP>
F;111-415/Product: carboxypeptidase B, pancreatic #status predicted <MAT>
F;173-186,258-272/Disulfide bonds: #status predicted
F;176,179,303/Binding site: zinc (His, Glu, His) #status predicted
F;355,377/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:                7.52e-48      Length:          416
Score:                    634.00        Matches:         141
Percent Similarity:       51.90%        Conservative:    78
Best Local Similarity:    33.41%        Mismatches:      141
Query Match:              22.83%        Indels:          63
DB:                       1             Gaps:            9

US-09-980-881A-1 (1-1573) x A42332 (1-416)

Qy          24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
               :::   :::|||   |||||||||   :::   |||           |||
Db           1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy          84 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
                  |||:::     :::|||                  :::   :::
Db          21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy         141 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 200
               ::::::   |||   :::|||   :::|||      |||:::|||   :::
Db          41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy         201 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 260
               |||               |||   |||   |||   |||   |||         |||:::
Db          61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy         261 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 320
                  |||      :::   :::         ||||||::::::::::::   ::::::
Db          81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy         321 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 380
               ::::::   |||      :::                |||||::         ||||||
Db         101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy         381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
               :::|||:::          |||   :::|||   :::   :::         :::|||
Db         118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy         441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
               ::::::::::::      ||||||::::::||||||         :::|||:::|||
Db         138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy         501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
               ||||||   ||||||            |||   ||||||::::::|||||||||   |||
Db         158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy         561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 602
               ||||||||||||||||||||||||||||||   ||||||:::
Db         177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy         602 ------------------------------------------------------------ 602

Db         197 TyrGlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPro 216

Qy         603 --------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629
                           |||:::                 :::|||   |||||||||
Db         217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy         630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
               ||||||   :::      :::      ||||||||||||   |||   ||||||
Db         237 ArgSerThrHisThrGlySer---SerIleGlyThrAspProAsnArgAsnPhe---Asp 254

Qy         690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
               |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db         255 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 274

Qy         750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
               |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db         275 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 294
```

```
Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
              ||||||||||||:::::::::|||||||||||||||   :::::::::|||||||||
Db        295 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 314

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
              |||                 :::::::::   ||||||:::   :::|||         |||
Db        315 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 334

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
              :::   :::      :::          |||:::||||||:::|||   |||:::   |||
Db        335 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 353

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
              :::|||   |||   ||||||   |||||||||   ||||||   ||||||:::||||||
Db        354 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 373

Qy       1050 TTTACATC---------------------------------------------------- 1057
              ||||||
Db        374 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 393

Qy       1058 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGG 1117
              :::   ||||||   |||   |||   |||:::      :::|||   :::|||:::
Db        394 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 413

Qy       1118 AATGTT 1123
              ::::::
Db        414 HisLeu 415

RESULT 6
CPBOB
carboxypeptidase B (EC 3.4.17.2) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
C;Accession: A93797; A92150; A00912
R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A;Title: Amino-acid sequence of bovine carboxypeptidase B.
A;Reference number: A93797; MUID:75217824; PMID:1057162
A;Accession: A93797
A;Molecule type: protein
A;Residues: 1-306 <TIT>
R;Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations o
A;Reference number: A92150; MUID:74260705; PMID:4833744
A;Accession: A92150
A;Molecule type: protein
A;Residues: 31-93;131-181;263-265;292-306 <SCH>
R;Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A;Reference number: A92839; MUID:76265065; PMID:957425
A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R;Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptid
A;Reference number: A92051; MUID:70007159; PMID:5344132
A;Contents: annotation; active site
R;Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptida
A;Reference number: A92119; MUID:73061487; PMID:4565668
A;Contents: annotation; active site
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-158,149-163/Disulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

Alignment Scores:
Pred. No.:              6.3e-44         Length:         306
Score:                  589.50          Matches:        121
Percent Similarity:     56.58%          Conservative:   51
Best Local Similarity:  39.80%          Mismatches:     75
Query Match:            21.23%          Indels:         58
DB:                     1               Gaps:           6

US-09-980-881A-1 (1-1573) x CPBOB (1-306)

Qy        375 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 434
              ||||||:::|||::::::          |||   :::|||   |||   :::
Db          6 TyrGluLysTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25

Qy        435 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 494
              :::||||||:::::::::::::      ||||||::::::|||           :::|||
Db         26 AsnProAspLeuIleSerArgSerAlaIleGlyThrThrPheLeuGlyAsnThrIleTyr 45

Qy        495 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 554
              :::|||||||||   ||||||      :::   |||   |||:::::::::|||||||||
Db         46 LeuLeuLysVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64

Qy        555 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 602
              ||||||||||||||||||||||||||||||   ||||||:::
Db         65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84

Qy        602 ------------------------------------------------------------ 602

Db         85 ArgThrTyrGlyArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104

Qy        603 --------------------GGCCAT-----------------AATCGAATGTGGAGA 623
                                  |||:::                 ||||||||||||||
Db        105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrThrAsnArgMetTrpArg 124

Qy        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683
              |||   ||||||      |||   :::   |||   ||||||||||||||   ||||||
Db        125 LysThrArgSerThrArgAlaGlySerSerCysThrGlyThrAspLeuAsnArgAsnPhe 144

Qy        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
                      ||||||      |||||||||::::::   ||||||||||||||||||
Db        145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163

Qy        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803
              |||         |||||||||   |||   ||||||||||||   |||:::|||   :::
Db        164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgAsnHis 183

Qy        804 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 863
              ::::::   ||||||||||||:::::::::|||||||||||||||   :::::::::|||
Db        184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203

Qy        864 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 923
              ||||||||                |||::::::   ||||||:::   :::|||
Db        204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223

Qy        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983
              ||||||:::   :::      :::          |||   |||:::::|||   |||:::
Db        224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242

Qy        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1043
              |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db        243 ThrThrIleTyrProAlaSerGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLys 262

Qy       1044 TATTCGTTTACATC---------------------------------------------- 1057
              ||||||||||||
Db        263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282

Qy       1058 ------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTC 1111
                    :::|||||||||   |||          |||:::      :::      :::|||
Db        283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIleLysTyrValThrSerTyrVal 302

Qy       1112 ATTAGGAATGTT 1123
              :::   ::::::
Db        303 LeuGluHisLeu 306

RESULT 7
A38395
```

```
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C;Accession: A38395; A33118
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A;Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme acti
A;Reference number: A38395; MUID:91105153; PMID:1988052
A;Accession: A38395
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-309 <COL>
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A;Reference number: A33118
A;Accession: A33118
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-309 <CO2>
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase
F;248,270/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:              5.59e-42      Length:          309
Score:                  567.50        Matches:         120
Percent Similarity:     54.79%        Conservative:    46
Best Local Similarity:  39.60%        Mismatches:      80
Query Match:            20.44%        Indels:          58
DB:                     2             Gaps:            6

US-09-980-881A-1 (1-1573) x A38395 (1-309)

Qy        375 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 434
              |||   :::|||:::     |||:::|||   ||||||  |||   :::   |||:::
Db          8 TyrAlaLysTyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLys 27

Qy        435 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 494
              ||||||:::|||:::::::::|||   ||||||||::   |||      |||||||||
Db         28 HisProGluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyr 47

Qy        495 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 554
              |||||||||:::   |||::::::         :::   ||||||:::::|||||||||
Db         48 ValLeuLysIle---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66

Qy        555 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 602
              ||||||||||||||||||:::|||||||||||||||   ||||||:::
Db         67 IleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86

Qy        603 ---------GGCCATAAT---------------------------------------- 611
                        |||   |||
Db         87 LysSerTyrGlyLysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106

Qy        612 ------------------------------------------------CGAATGTGGAGA 623
                                                             ||||||||||||
Db        107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126

Qy        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683
              ||||||||||||        ::::::   ||||||||||||||||||||   ||||||
Db        127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146

Qy        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
                      |||               :::   :::   |||          |||
Db        147 ---AspValSerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArg 165

Qy        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803
              |||      ||||||||||||   |||   |||||||||   :::|||:::|||   :::
Db        166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185

Qy        804 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 863
              :::|||   ||||||||||||||:::   ||||||||||||||   ::::::||||||
Db        186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205

Qy        864 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 923
              |||   ||||||              :::|||::::::|||      ||||||
Db        206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIle 225

Qy        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983
              |||         :::      :::         |||||||||   :::|||      :::
Db        226 AlaThrAspValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAla 244

Qy        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1043
              |||:::|||         |||         ||||||   ||||||||||||||||||
Db        245 SerThrIleTyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLys 264

Qy       1044 TATTCGTTTAC---------------------------------------------------- 1054
              ::::::|||
Db        265 HisThrPheAlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSer 284

Qy       1055 ---ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTC 1111
                 ||||||||||||||:::|||         :::|||      ||||||   ::::::
Db        285 ArgIleLysProThrCysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIle 304

Qy       1112 ATTAGGAAT 1120
              :::::::::
Db        305 LeuLysHis 307

RESULT 8
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C;Accession: A32128
R;Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rutt
J. Biol. Chem. 263, 17828-17836, 1988
A;Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evo
A;Reference number: A32128; MUID:89034323; PMID:3182871
A;Accession: A32128
A;Molecule type: mRNA
A;Residues: 1-417 <GAR>
A;Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g5
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:              1.11e-40      Length:          417
Score:                  553.00        Matches:         137
Percent Similarity:     50.37%        Conservative:    69
Best Local Similarity:  33.50%        Mismatches:      137
Query Match:            19.91%        Indels:          66
DB:                     2             Gaps:            13

US-09-980-881A-1 (1-1573) x A32128 (1-417)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||:::|||      |||         :::         ::::::|||::::::
Db          1 MetArgLeuThrLeuLeuLeuAlaAlaLeuLeuGlyTyrIleTyrCysGlnGlu------ 18

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
                 |||         |||||||||      :::|||         |||::::::
Db         19 ---ThrPheValGlyAspGlnValLeuGluIleIleProSerHisGluGluGlnIleArg 37

Qy        138 GTTCTACAGAATCTTACTACAACA------TATGAGATTGTTCTCTGGCAGCCGGTAACA 191
                 |||      |||                  |||:::      |||:::      |||
Db         38 ThrLeuLeuGlnLeuGluAlaGluGluHisLeuGluLeuAspPheTrpLysSerProThr 57

Qy        192 GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGAC 251
                       :::      :::   ||||||      |||      :::   :::
Db         58 ------------IleProGlyGluThrValHisValArgValProPheAlaSerIleGln 73

Qy        252 AATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTG 311
                 ||||||      |||         ||||||      |||::::::::::   ||||||
Db         74 AlaValLysValPheLeuGluSerGlnGlyIleAspTyrSerIleMetIleGluAspVal 93

Qy        312 GAAGATCTTATTCAACAGCAGATTTCCAAC---------GACACAGTCAGCCCCCGAGCC 362
```

```
Db         94 GlnValLeuLeuAspGlnGluArgGluGluMetLeuPheAsnGlnGlnArgGluArgGly 113

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
Db        114 GlyAsnPheAsnPheGluAlaTyrHisThrLeuGluGluIleTyrGlnGluMetAspAsn 133

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
Db        134 LeuValAlaGluAsnProGlyLeuValSerLysValAsnLeuGlySerSerPheGluAsn 153

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACAGCCAAAAATGCCATA 539
Db        154 ArgProMetAsnValLeuLysPheSerThrGlyGlyAsp---------LysProAlaIle 170

Qy        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGG--- 596
Db        171 TrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAlaThrAlaLeuTrpThr 190

Qy        596 ------------------------------------------------------------ 596

Db        191 AlaAsnLysIleAlaSerAspTyrGlyThrAspProAlaIleThrSerLeuLeuAsnThr 210

Qy        597 ---------TTCATA---------------------GGCCAT------------------ 608
Db        211 LeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrValPheSerGlnThrThr 230

Qy        609 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 668
Db        231 AsnArgMetTrpArgLysThrArgSerLysArgSerGlySerGlyCysValGlyValAsp 250

Qy        669 CTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 728
Db        251 ProAsnArgAsnTrp---AspAlaAsnPheGlyGlyProGlyAlaSerSerSerProCys 269

Qy        729 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 788
Db        270 SerAspSerTyrHisGlyProLysProAsnSerGluValGluValLysSerIleValAsp 289

Qy        789 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 848
Db        290 PheIleLysSerHis---GlyLysValLysAlaPheIleThrLeuHisSerTyrSerGln 308

Qy        849 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 908
Db        309 LeuLeuMetPheProTyrGlyTyrLysCysThrLysProAspAspPheAsnGluLeuAsp 328

Qy        909 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACA 968
Db        329 GluValAlaGlnLysAlaAlaGlnAlaLeu---LysArgLeuHisGlyThrSerTyrLys 347

Qy        969 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1028
Db        348 ValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySerIleAspTrpAlaTyr 367

Qy       1029 GATTTGGGCATCAAATATTCGTTTACATCAAAC-------------------------- 1061
Db        368 AspLeuGlyIleLysTyrSerPheAlaPheGluLeuArgAspThrAlaPheTyrGlyPhe 387

Qy       1062 ---CCACCTGTAGAGAAGCTTTTGCCG 1085
Db        388 LeuLeuProAlaLysGlnIleLeuPro 396
```

RESULT 9
CPBOA
carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine
N;Alternate names: procarboxypeptidase A activation peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1992 #text_change 05-May-2000
C;Accession: JN0126; JT0440; A31406; S00059; A38834; A00910; A90355
R;Le Hueerou, I.; Guilloteau, P.; Toullec, R.; Puigserver, A.; Wicker, C.
Biochem. Biophys. Res. Commun. 175, 110-116, 1991
A;Title: Cloning and nucleotide sequence of a bovine pancreatic preprocarboxypeptidase A
A;Reference number: JN0126; MUID:91151335; PMID:1998496
A;Accession: JN0126
A;Molecule type: mRNA
A;Residues: 1-419 <LEH>
A;Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R;Foglizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A;Title: Primary structure of the activation peptide from bovine pancreatic procarboxypep
A;Reference number: JT0440; MUID:88209583; PMID:3365436
A;Accession: JT0440
A;Molecule type: protein
A;Residues: 17-110 <FOG>
R;Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A;Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidase
A;Reference number: A31406; MUID:89150306; PMID:3147705
A;Accession: A31406
A;Molecule type: protein
A;Residues: 17-110 <WAD>
R;Chapus, C.; Kerfelec, B.; Foglizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A;Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A;Reference number: S00059; MUID:87275921; PMID:3609014
A;Accession: S00059
A;Molecule type: protein
A;Residues: 17-69 <CHA>
R;Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A;Title: The amino acid sequence of bovine carboxypeptidase A.
A;Reference number: A38834; MUID:70050818; PMID:5260942
A;Accession: A38834
A;Molecule type: protein
A;Residues: 111-137,'Q',139-140,'E',142-417 <BRA>
R;Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A;Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterizatio
A;Reference number: A90360; MUID:71155185; PMID:4927805
A;Accession: A00910
A;Molecule type: protein
A;Residues: 214-411 <BR2>
A;Note: this is the final paper in a series
R;Petra, P.H.; Hermodson, M.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 4023-4025, 1971
A;Title: Characterization of bovine carboxypeptidase A(Allan).
A;Reference number: A90355; MUID:72138789; PMID:5143102
A;Accession: A90355
A;Molecule type: protein
A;Residues: 111-142 <PET>
R;Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A;Title: Identification of the amino acid replacements characterizing the allotropic form
A;Reference number: A90556; MUID:69283620; PMID:5817619
A;Contents: annotation; allelic variant
A;Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CPA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137,'Q',139-14
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/Disulfide bonds: #status experimental
F;358,380/Active site: Tyr, Glu #status experimental

Alignment Scores:

Pred. No.:	3.56e-39	Length:	419
Score:	536.00	Matches:	141
Percent Similarity:	47.80%	Conservative:	65
Best Local Similarity:	32.71%	Mismatches:	145
Query Match:	19.30%	Indels:	81
DB:	1	Gaps:	12

US-09-980-881A-1 (1-1573) x CPBOA (1-419)

```
Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
              |||      |||:::||||||:::         |||      |||
Db          4 LeuLeuIleLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp------------ 19

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
              |||    . |||||||||                      :::||||||   :::
Db         20 PheValGlyHisGlnValLeuArgIleThrAlaAlaAspGluAlaGluValGlnThrVal 39

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGG------------CAGCCG 185
              :::   |||                :::   |||      |||             ||||||
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProGlyGlnPro 59

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                                          :::      |||
Db         60 ------------------------GlySerProIleAspValArgValProPheProSer 71

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
              :::      ||||||      |||         ||||||         :::::::::
Db         72 LeuGlnAlaValLysValPheLeuGluAlaHisGlyIleArgTyrArgIleMetIleGlu 91

Qy        306 GACGTGGAAGATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCC 356
              ||||||:::   |||:::         |||:::|||:::
Db         92 AspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaSerGlnSerArgAla 111

Qy        357 CGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 413
              |||::::::   ::::::   |||      ||||||:::|||:::|||||||||   :::
Db        112 ArgSerThrAsnThrPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrAspPhe 131

Qy        414 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 473
              ::::::   :::         ||||||   :::::::::|||:::   ||||||   |||
Db        132 MetAspLeuLeuValAlaGluHisProGlnLeuValSerLysLeuGlnIleGlyArgSer 151

Qy        474 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 533
              :::|||      |||:::||||||||||||   |||              :::   :::
Db        152 TyrGluGlyArgProIleTyrValLeuLysPheSer---ThrGlyGlySerAsnArgPro 170

Qy        534 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 593
              |||||||||||||||   |||||||||:::|||||||||||:::   |||      :::
Db        171 AlaIleTrpIleAspLeuGlyIleHisSerArgGluTrpIleThrGlnAlaThrGlyVal 190

Qy        594 TGGTTCATA---------------------------------------------------- 602
              ||||||
Db        191 TrpPheAlaLysLysPheThrGluAspTyrGlyGlnAspProSerPheThrAlaIleLeu 210

Qy        602 ------------------------------------------------------------ 602

Db        211 AspSerMetAspIlePheLeuGluIleValThrAsnProAspGlyPheAlaPheThrHis 230

Qy        603 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 662
              ||||||:::|||||||||   ||||||      :::::::::   |||:::|||
Db        231 SerGlnAsnArgLeuTrpArgLysThrArgSerValThrSerSerSerLeuCysValGly 250

Qy        663 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 722
                 |||   |||   |||:::            :::   :::   ||||||||||||||
Db        251 ValAspAlaAsnArgAsnTrp---AspAlaGlyPheGlyLysAlaGlyAlaSerSerSer 269

Qy        723 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 782
                 ||||||||||||||   |||   |||      ||||||   |||||||||::::::
Db        270 ProCysSerGluThrTyrHisGlyLysTyrAlaAsnSerGluValGluValLysSerIle 289

Qy        783 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 842
                 |||::::::   :::   |||      ||||||::::::|||:::|||||||||
Db        290 ValAspPheValLysAspHisGlyAsn---PheLysAlaPheLeuSerIleHisSerTyr 308
```

```
Qy        843 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 902
              ||||||   :::::::::||||||   ||||||                |||      |||
Db        309 SerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrGlnSerIleProAspLysThrGlu 328

Qy        903 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 962
              |||:::   ||||||      ||||||   |||:::   |||:::         |||
Db        329 LeuAsnGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyrGlyThrSer 347

Qy        963 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1022
              |||   :::|||             |||:::|||   |||   ||||||      ||||||
Db        348 TyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSerGlyGlySerIleAspTrp 367

Qy       1023 ATCTATGATTTGGGCATCAAATATTCGTTTACATCAAA---------------------- 1060
                 |||:::   ||||||||||||||||||||
Db        368 SerTyrAsnGlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyArgTyr 387

Qy       1061 -----------------------------CCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1090
                                           ||||||   :::|||   :::      |||
Db        388 GlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGlnGluThrTrpLeuGlyVal 407

Qy       1091 TCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123
                 |||      |||   :::   |||:::
Db        408 LeuThrIleMetGluHisThrLeuAsnAsnLeu 418
```

RESULT 10
CPRTA
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.L
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <QUI>
A;Cross-references: GB:J00713; NID:g203364; PIDN:AAA40893.1; PID:g203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparat
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA
A;Residues: 1-260,'F',262,'M',264-346,'K',348-419 <CLA>
C;Genetics:
A;Gene: CPA1
A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

Alignment Scores:			
Pred. No.:	1.48e-38	Length:	419
Score:	529.00	Matches:	141
Percent Similarity:	47.94%	Conservative:	68
Best Local Similarity:	32.34%	Mismatches:	151
Query Match:	19.05%	Indels:	76
DB:	1	Gaps:	13

US-09-980-881A-1 (1-1573) x CPRTA (1-419)

```
Qy         21 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 80
              :::|||      |||::::::|||:::      |||      |||
Db          3 ArgLeuLeuIleLeuSerLeuLeuLeuGluAlaValCysGlyAsnGluAsn--------- 19

Qy         81 GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
```

```
Db         20 ---PheValGlyHisGlnValLeuArgIleSerAlaAlaAspGluAlaGlnValGlnLys 38

Qy        141 CTACAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAGCCGGTAACAGCT 194
Db         39 ValLysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg 58

Qy        195 GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT 254
Db         59 AlaGlyIleProIleAspValArgValProPhe-----------ProSerIleGlnSer 74

Qy        255 GTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAA 314
Db         75 ValLysAlaPheLeuGluTyrHisGlyIleSerTyrGluIleMetIleGluAspValGln 94

Qy        315 GATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCC 365
Db         95 LeuLeuLeuAspGluGluLysGlnGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr 114

Qy        366 GCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
Db        115 AspSerPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrGluPheMetAspLeu 134

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
Db        135 LeuValAlaGluHisProGlnLeuValSerLysIleGlnIleGlyAsnThrPheGluGly 154

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
Db        155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTrp 173

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
Db        174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193

Qy        603 GGCCAT---------------------------------------------------- 608

Db        194 LysLysValThrLysAspTyrGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213

Qy        609 ----------------------------------------------------------AAT 611
Db        214 AspIlePheLeuGluIleValThrAsnProAspGlyPheAlaTyrThrHisLysThrAsn 233

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
Db        234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuCysValGlyValAspPro 253

Qy        672 AATAGCAACTTTGTCTCCAAACACTGG-----------TGTGAGGAAGGTGCATCCAGT 719
Db        254 AsnArgAsn--------------TrpAspAlaGlyLeuGlyLysAlaGlyAlaSerSer 268

Qy        720 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 779
Db        269 AsnProCysSerGluThrTyrArgGlyLysPheProAsnSerGluValGluValLysSer 288

Qy        780 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 839
Db        289 IleValAspPheValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSer 307

Qy        840 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 899
Db        308 TyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrSerGluProAlaProAspGlnAla 327

Qy        900 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959
Db        328 GluLeuAspGlnLeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThr 346

Qy        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019
Db        347 GluPheLysTyrGlySerIleIleAspThrIleTyrGlnAlaSerGlySerThrIleAsp 366

Qy       1020 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC----------------- 1061
```

```
Db        367 TrpThrTyrSerGlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyLeu 386

Qy       1062 ------------CCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATG 1109
Db        387 ArgGlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGluGlu---------Thr 403

Qy       1110 TCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
Db        404 TrpLeuAlaLeuLeuThrIleMetAspHisThrValLysHisProTyr 419
```

RESULT 11
A56171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C;Accession: A56171; S02809; S71395
R;Catasus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA c
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAA74425.1; PID:g790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S',18-36,'N',38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton,
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LAE>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	1.07e-36	Length:	417
Score:	508.00	Matches:	128
Percent Similarity:	48.48%	Conservative:	63
Best Local Similarity:	32.49%	Mismatches:	137
Query Match:	18.29%	Indels:	66
DB:	2	Gaps:	11

US-09-980-881A-1 (1-1573) x A56171 (1-417)

```
Qy         72 CATGTCTTC---------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACC 122
Db         13 HisIleTyrCysLeuGluThrPheValGlyAspGlnValLeuGluIleValProSerAsn 32

Qy        123 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC------ 176
Db         33 GluGluGlnIleLysAsnLeuLeuGlnLeuGluAlaGlnGluHisLeuGlnLeuAspPhe 52
```

```
Qy        177 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 236
              |||:::     |||                 :::      |||      |||
Db         53 TrpLysSerProThr-----------ThrProGlyGluThrAlaHisValArgValPro 68

Qy        237 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 296
                 :::|||      ||||||      |||       ||||||      |||:::
Db         69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTyrSerIle 88

Qy        297 TTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCC 356
              ::::::   ||||||:::   |||:::   ::::::                  :::
Db         89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108

Qy        357 CGAGCCTCCGCATCG---------TACTATGAACAGTATCACTCACTAAATGAAATCTAT 407
              |||         |||           :::      ||||||:::|||   ||||||
Db        109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyrHisThrLeuGluGluIleSer 128

Qy        408 TCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGA 467
                 ::::::   :::         ||||||   :::::::::|||::::::||||||
Db        129 GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148

Qy        468 TCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACA 524
              ||||||||||||      |||:::   |||||||||   |||   |||   :::
Db        149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp------ 166

Qy        525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
                 |||   |||||||||:::|||   ||||||||||||||||||||::::::   |||
Db        167 ---LysProAlaIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185

Qy        585 TTCTGCTTGTGGTTCATAGGC---------------------------------------- 605
                 ||||||
Db        186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205

Qy        605 ------------------------------------------------------------ 605

Db        206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225

Qy        606 ------------CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 653
                          ||||||||||||||||||   ||||||      :::   :::
Db        226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu 245

Qy        654 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 713
              |||:::|||   |||   |||   |||:::           :::         ||||||
Db        246 CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264

Qy        714 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 773
              ||||||:::   ||||||:::::|||   |||             ||||||   ||||||
Db        265 SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal 284

Qy        774 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 833
              |||::::::      |||::::::   :::      ::::::||||||:::|||   :::
Db        285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleIleLeu 303

Qy        834 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 893
              ||||||||||||||   ::::::||||||||   |||       :::|||       |||
Db        304 HisSerTyrSerGlnLeuLeuMetPheProTyrGlyTyrLysCysThrLysLeuAspAsp 323

Qy        894 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 953
                 :::|||||||||   ||||||   :::|||   :::::::::   ::::::
Db        324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342

Qy        954 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 1013
              |||:::|||      |||                 :::|||   |||   ||||||
Db        343 GlyThrLysTyrLysValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer 362

Qy       1014 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC------------ 1061
              ||||||   ||||||   ||||||||||||||||||
Db        363 IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr 382

Qy       1062 -----------------CCACCTGTAGAGAAGCTTTTGCCG 1085
                               |||      ::::::||||||
Db        383 GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro 396

RESULT 12
S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N;Alternate names: pancreatic carboxypeptidase A1
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: S29127; A34205; S08253; S02810; S71394; S02811
R;Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver,
Biochem. J. 287, 299-303, 1992
A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.
A;Reference number: S29127; MUID:93038569; PMID:1417781
A;Accession: S29127
A;Molecule type: mRNA
A;Residues: 1-419 <CAT>
A;Cross-references: EMBL:X67318; NID:g35329; PIDN:CAA47732.1; PID:g35330
R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A;Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 330-396 <STE>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxypep
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43;'XXX',114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton,
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:                4.66e-35        Length:          419
Score:                    489.50          Matches:         131
Percent Similarity:       47.63%          Conservative:    60
Best Local Similarity:    32.67%          Mismatches:      127
Query Match:              17.63%          Indels:          83
DB:                       1               Gaps:            13

US-09-980-881A-1 (1-1573) x S29127 (1-419)

Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
              |||      |||:::||||||:::      |||       |||
Db          4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluAsp------------ 19

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
```

```
Db         20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValGlnLysVal 39

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAG---------------- 182
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59

Qy        183 ------CCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 236
Db         60 GlySerProIleAspVal--------------------ArgValProPhe--------- 69

Qy        237 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 296
Db         70 ---ProSerIleGlnAlaValLysIlePheLeuGluSerHisGlyIleSerTyrGluThr 88

Qy        297 TTGCTGGCAGACGTGGAAGATCTTATT---------CAACAGCAGATT------------ 335
Db         89 MetIleGluAspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaPheArg 108

Qy        336 ------------TCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
Db        109 SerArgAlaArgSerThrAspThrPheAsn--------------------TyrAlaThr 121

Qy        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
Db        122 TyrHisThrLeuGluGluIleTyrAspPheLeuAspLeuLeuValAlaGluAsnProHis 141

Qy        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
Db        142 LeuValSerLysIleGlnIleGlyAsnThrTyrGluGlyArgProIleTyrValLeuLys 161

Qy        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
Db        162 PheSer---ThrGlyGlySerLysArgProAlaIleTrpIleAspThrGlyIleHisSer 180

Qy        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
Db        181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyr 200

Qy        602 ------------------------------------------------------------ 602
Db        201 GlyGlnAspAlaAlaPheThrAlaIleLeuAspThrLeuAspIlePheLeuGluIleVal 220

Qy        603 ----------------------------GGCCATAATCGAATGTGGAGAAAGAACCGT 632
Db        221 ThrAsnProAspGlyPheAlaPheThrHisSerThrAsnArgMetTrpArgLysThrArg 240

Qy        633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
Db        241 SerHisThrAlaGlySerLeuCysIleGlyValAspProAsnArgAsnTrp---AspAla 259

Qy        693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
Db        260 GlyPheGlyLeuSerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyLysPhe 279

Qy        753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
Db        280 AlaAsnSerGluValGluValLysSerIleValAspPheValLysAspHisGlyAsn--- 298

Qy        813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
Db        299 IleLysAlaPheIleSerIleHisSerTyrSerGlnLeuLeuMetTyrProTyrGlyTyr 318

Qy        873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
Db        319 LysThrGluProValProAspGlnAspGluLeuAspGlnLeuSerLysAlaAlaValThr 338

Qy        933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
Db        339 AlaLeu---AlaSerLeuTyrGlyThrLysPheAsnTyrGlySerIleIleLysAlaIle 357

Qy        993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
```

```
Db        358 TyrGlnAlaSerGlySerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377

Qy       1053 ACA 1055
Db        378 Thr 378
```

RESULT 13
T33527
hypothetical protein T06A4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C;Accession: T33527
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33527
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-528 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.1
A;Map position: 1
A;Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C;Superfamily: carboxypeptidase

Alignment Scores:

Pred. No.:	1.92e-30	Length:	528
Score:	437.50	Matches:	129
Percent Similarity:	45.28%	Conservative:	63
Best Local Similarity:	30.42%	Mismatches:	139
Query Match:	15.75%	Indels:	93
DB:	2	Gaps:	16

US-09-980-881A-1 (1-1573) x T33527 (1-528)

```
Qy         18 ATGAAGCTTTGCAGC---CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCAT 74
Db          1 MetThrLysCysGlnIleLeuAlaSerLeuLeuThrTrpSerPheTrpValPheArgHis 20

Qy         75 GTCTTCGCGTTCCAGAGTGGC---------------------------------------- 95
Db         21 ValLeuAlaThrThrAspGlyArgAlaPheArgLysLysSerSerAlaThrThrProPro 40

Qy         96 ----------------------CAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTT 134
Db         41 AsnGlnLysThrSerSerPheLysLeuIleArgIleAsnProGluThrGluGlySerVal 60

Qy        135 CAAGTTCTACAGAATCTT---------ACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
Db         61 LysTyrLeuArgSerLeuTyrGluAspProSerProTyrGluLeuAspPheTrpGlnPro 80

Qy        186 GTAACA--------------GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTT 230
Db         81 ProThrAsnIleGlyAlaIleValAspLeuThrValAlaProAlaAspAlaProArgPhe 100

Qy        231 GTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 290
Db        101 Val------LysAspLeuGluSerLysLys--------------------IleSerTyr 111

Qy        291 AGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 350
Db        112 IleValAlaValAsnAspLeuSerLysAlaIleGluAsnGluArgGlySerAspLysPhe 131

Qy        351 AGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCT 410
Db        132 TyrAsnProValAlaGlyPheAlaTyrAspLysTyrAsnSerLeuGluGluIleGlnThr 151

Qy        411 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 470
Db        152 GluMetLysArgLeuLysLysGluTyrProThrMetIleThrLeuIleAspIleGlyGln 171
```

```
Qy        471 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA------ 524
Db        172 SerHisGluAsnArgThrLeuLeuValMetLysIleThrGlyLysArgAsnProLeuGly 191

Qy        525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
Db        192 SerLysIleSerMetTrpIleAspAlaGlyIleHisAlaArgGluTrpIleAlaProAla 211

Qy        585 TTCTGCTTGTGGTTCATAGGCCAT---------------------------------- 608
Db        212 ThrAlaMet---TyrIleAlaHisGluLeuLeuLeuGlyTyrGluAsnAspAlaThrVal 230

Qy        608 ------------------------------------------------------------ 608

Db        231 AlaLysLeuMetAspHisIleAspPheTyrIleLeuProValMetAsnProAspGlyTyr 250

Qy        609 ----------AATCGAATGTGGAGAAAGAACCGTTCT-------------- 635
Db        251 GluTyrSerArgGluLysAsnArgMetTrpArgLysAsnArgSerProAlaLysCysAla 270

Qy        636 ------TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
Db        271 ArgGlnThrPheSerThrValCysCysSerGlyValAspLeuAsnArgAsnPhe---Asp 289

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
Db        290 TrpPheTrpAlaSerThrGlySerSerSerAspProCysHisAspThrTyrHisGlySer 309

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
Db        310 AlaAlaPheSerGluProGluSerGlnAlaValArgAspPheLeuGluGlnAsnThrPro 329

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
Db        330 Glu------AlaPheIleSerLeuHisSerTyrSerGlnMetTrpLeuIleProTyrGly 347

Qy        870 TATACACGA---AGTAAAAGCAAAGAC---CATGAGGAACTGTCTCTAGTAGCCAGTGAA 923
Db        348 HisArgLysGlnSerTyrProGlnAspTyrHisThrGlyLeuArgProLeuAlaLeuArg 367

Qy        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983
Db        368 AlaThrLysAlaLeuTyrGluLeu---TyrGlyThrLysTyrGlnValGlyThrGlyAla 386

Qy        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT---GATTTGGGCATC 1040
Db        387 AspLeuMetTyrGluAlaSerGlyGlySerHisAspTrpAlaLysGlyGlnLeuLysVal 406

Qy       1041 AAATATTCGTTT 1052
Db        407 ProTyrAlaTyr 410
```

RESULT 14
T33526
hypothetical protein T06A4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33526
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-667 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67474.1; GSPDB:GN00019; CESP:T06A4.3
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.3
A;Map position: 1
A;Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

Alignment Scores:

Pred. No.:	3.73e-27	Length:	667
Score:	400.50	Matches:	131
Percent Similarity:	40.70%	Conservative:	66
Best Local Similarity:	27.07%	Mismatches:	148
Query Match:	14.42%	Indels:	139
DB:	2	Gaps:	16

US-09-980-881A-1 (1-1573) x T33526 (1-667)

```
Qy         51 ATTGTTCTCTTCTGTGAGCAGCATGTC---TTCGCGTTCCAGAGTGGCCAAGTTCTAGCT 107
Db          8 LeuIlePheLeuCysPheSerAlaValSerLeuSerThrGlnSerPheAlaValTyrArg 27

Qy        108 GCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT---------ACTACA 158
Db         28 LeuLeuProLysSerGlnThrAspPheGlnAlaIleGlnArgLeuTyrLysAsnAlaThr 47

Qy        159 ACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCT--------------GACCTTATT 203
Db         48 AspHisAspLeuAsnPheTrpLysThrGlyLysAspLysHisGlyPheTrpAspValMet 67

Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
Db         68 ValAspMetLysAsnSerLysTyrPheLeuAsp------------------------- 78

Qy        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
Db         79 PheLeuGlnValAsnAspIleSerTyrIleLysThrIleAspAspValGluGlyLeuIle 98

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGC---------CCCCGA--------------- 359
Db         99 ThrLysHisGluAsnAsnLysThrLeuSerAsnMetProArgLeuTrpAspAspSer 118

Qy        360 GCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATA 416
Db        119 SerSerAlaHisTyrAspPheHisThrTyrGlySerTyrGlnArgMetThrAspTrpMet 138

Qy        417 GAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTT 476
Db        139 LysGlnLeuValValLysTyrProLysMetValGlnTyrIleSerIleGlyLysThrThr 158

Qy        477 GAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCC 536
Db        159 GluGlyArgAsnIleAspGlyValGluIleGlyGlyAspSerArgThr---LysLysIle 177

Qy        537 ATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGG 596
Db        178 PheTrpIleAspGlyGlyIleHisAlaArgGluTrpAlaAlaProHisThrAlaLeuPhe 197

Qy        597 TTCATA---------------------------------------------------- 602
Db        198 PheIleHisGlnValCysAspProAlaProSerGluLysSerValPheArgLeuValPro 217

Qy        602 ------------------------------------------------------------ 602

Db        218 LeuGlyLeuHisLysCysTrpGlnArgGlnLysLysGlyLysLeuAspAsnLysTrpLeu 237

Qy        602 ------------------------------------------------------------ 602

Db        238 ThrSerArgAlaAsnGluProGlyIleLysLysLeuLeuAsnGluIleThrPheValVal 257

Qy        603 ----------------------------------------------------GGCCATAAT 611
Db        258 ValProCysLeuAsnProAspGlyTyrGluPheThrArgSerSerThrAsnProHisVal 277

Qy        612 CGAATGTGGAGAAAGAACCGTTCT----------------------TTCTATGCGAAC 647
Db        278 ArgLeuTrpArgLysAsnArgSerLysMetGlnCysArgLysAspIleTrpGlyArgAsn 297

Qy        648 AATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAA 707
```

```
Db        298 ArgCysCysArgGlyValAspLeuAsnArgAsnPhe---AspPheHisPheArgGluSer 316

Qy        708 GGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCA 767
              |||   |||         |||||||||   |||   |||      |||   |||||||||
Db        317 GlyThrSerAspAspProCysSerGluIleTyrGlnGlyProSerProPheSerGluPro 336

Qy        768 GAAGTGAAGGCAGTGGCTAGTTTCTTG------AGAAGAAATATCAACCAGATTAAAGCA 821
              |||   |||||||||         |||      :::|||         :::      |||
Db        337 GluAlaLysAlaValArgAspAlaLeuLeuSerGlnArgTyrLysGlyArgThrAspAla 356

Qy        822 TACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA--- 878
              ||||||::::::|||:::|||||||||     :::   ||||||   :::   :::
Db        357 TyrIleThrLeuHisThrTyrSerGlnIleTrpIleHisProTyrGlyHisLysLysAsp 376

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
              :::         |||   ::::::|||      |||      :::|||   :::|||:::
Db        377 AlaTyrProGlyAspIleLysAspLeuTyrGluValGlyLysLysAlaAlaGlnAlaLeu 396

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
                 |||            |||:::|||      |||   |||::::::|||||||||
Db        397 ---LysArgValTyrGlyThrLysTyrValValGlySerGlyAlaAspThrLeuTyrPro 415

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATC---TATGATTTGGGCATCAAATATTCGTTTACA 1055
              |||   ||||||   :::||||||      ::::::      :::|||  |||||||||:
Db        416 AlaSerGlyGlySerGluAspTrpAlaLysHisGluAlaLysValLys--PheValTyrL 435

Qy       1056 TCAAA---------------------------------------------------------- 1060
              ::
Db        435 euLeuGluLeuArgProAspGluLysAsnTrpAspGlyPheIleLeuAspGluLysGluL 455

Qy       1061 -----CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTA 1115
                   ||||||   ||||||   :::      |||      :::|||      |||:::
Db        455 euIleProThrAlaArgGluThrTrpGluGlyValArgValValAlaGluAlaValLeuA 475

Qy       1116 GGAATGTT 1123
                 :::
Db        475 spArgIle 477

RESULT 15
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20507; T24289
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284
A;Accession: T20507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WIL>
A;Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone F02D8
R;Wild, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19869
A;Accession: T24289
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WI2>
A;Cross-references: EMBL:Z78413; PIDN:CAB01666.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone T01C3
C;Genetics:
A;Gene: CESP:F02D8.4
A;Map position: 5
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C;Superfamily: carboxypeptidase

Alignment Scores:
Pred. No.:              6.97e-27        Length:         323
Score:                  397.00          Matches:        88
Percent Similarity:     53.01%          Conservative:   53
Best Local Similarity:  33.08%          Mismatches:     81

Query Match:            14.30%          Indels:         44
DB:                     2               Gaps:           7

US-09-980-881A-1 (1-1573) x T20507 (1-323)

Qy        411 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 470
              ::::::      :::   :::::::::|||:::   :::         :::||||||:::
Db          3 PheLeuAsnSerLeuAlaGlnGlnTyrProAsnAspValLysLeuGlnAsnIleGlyAsn 22

Qy        471 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAA 530
              |||:::|||         :::      ::::::::::      :::   ::::::|||
Db         23 SerTyrGluGlyArgSerIleThrAlaValArgIleAla---AspAspGlySerSerLys 41

Qy        531 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGC 590
                  :::|||||||||   ||||||||||||||||||||||||||
Db         42 ProIleValTrpIleAspAlaGlyIleHisAlaArgGluTrpIleSerTyrAsnValAla 61

Qy        591 TTGTGGTTCATA---------------------------------------------- 602
              |||:::   |||
Db         62 LeuTyrLeuIleTyrThrIleValSerGlnProAlaTyrArgAsnLeuLeuAspSerVal 81

Qy        603 ----------------------------------------------GGCCATAAT 611
                                                           ::::::
Db         82 GlnLeuValValValProAsnThrAsnProAspGlyTyrGluTyrSerArgThrAsnAsp 101

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              |||||||||||||||   ||||||   :::   |||:::   |||   |||   |||
Db        102 ArgMetTrpArgLysThrArgSerArgPheThrAsnSerArgCysAlaGlyAlaAspAla 121

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   |||:::         :::|||      :::|||   |||   |||   ||||||
Db        122 AsnArgAsnTyr---ProPheTyrTrpGlyThrGlnGlyValSerHisSerGlnCysSer 140

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              |||   :::||||||      |||:::||||||||||||||   |||:::   :::
Db        141 GluIlePheCysGlySerArgProGlnSerGluProGluValLeuAlaLeuThrAsnAla 160

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::   |||:::      :::||||||   ||||||::::::|||||||||   |||
Db        161 IleIleArgAspGluGluArgIleLysGlyTyrIleAlaLeuHisSerTyrGlyGlnGlu 180

Qy        852 ATAGTGTTTCCATATTCCTATACA---CGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 908
              |||::::::|||:::   :::|||   |||:::         |||   ::::::|||
Db        181 IleLeuTyrProTrpGlyHisThrGlnArgThrTyrProThrAspValGlnAspLeuIle 200

Qy        909 CTAGTAGCC---------AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959
                 |||            :::   |||:::||||||::::::            ||||||
Db        201 GlnValGlyArgAlaMetAlaSerAlaIleArgAlaValAsn-----------AsnThr 216

Qy        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019
                 ||||||      :::      :::   ||||||   |||   |||      ||||||
Db        217 AspTyrThrValValAsnSerGlyAspGlyLeuTyrProAlaAlaGlyAlaSerAspAsp 236

Qy       1020 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC---CCACCTGTAGAGAAG 1076
              |||            |||||||||||||:::|||            |||::::::
Db        237 TrpAlaLysSerArgGlyIleLysTyrSerTyrThrIleGluLeuSerProIleAspAsp 256

Qy       1077 CTTTTGCCGCTGTCTCTA 1094
                        ||||||
Db        257 PheThrGlyPheSerLeu 262


Search completed: January  1, 2004, 19:00:31
Job time : 57.8315 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:41:10 ; Search time 16.6016 Seconds
(without alignments)
8911.561 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135748_24941/app_query.fasta_1.2958
-DB=SwissProt_41 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_34_@runat_31122003_135748_24941 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	695.5	25.0	415	1	CBPB_RAT	P19223 rattus norv
2	655	23.6	417	1	CBPC_MOUSE	P15089 mus musculu
3	652.5	23.5	417	1	CBPB_HUMAN	P15086 homo sapien
4	647.5	23.3	417	1	CBPC_HUMAN	P15088 homo sapien
5	640.5	23.1	416	1	CBPB_CANFA	P55261 canis famil
6	629.5	22.7	401	1	CBPB_PIG	P09955 sus scrofa
7	589.5	21.2	306	1	CBPB_BOVIN	P00732 bos taurus
8	567.5	20.4	309	1	CBPC_RAT	P21961 rattus norv
9	553	19.9	417	1	CBP2_RAT	P19222 rattus norv
10	536	19.3	419	1	CBPA_BOVIN	P00730 bos taurus
11	530	19.1	419	1	CBP1_RAT	P00731 rattus norv
12	511	18.4	417	1	CPB2_HUMAN	P48052 homo sapien
13	489.5	17.6	419	1	CBP1_HUMAN	P15085 homo sapien
14	483	17.4	421	1	CBP4_HUMAN	Q9ui42 homo sapien
15	409	14.7	304	1	CBPZ_SIMVI	P42788 simulium vi
16	386	13.9	433	1	CBPA_ANOGA	O02350 anopheles g
17	369.5	13.3	303	1	CBPB_ASTFL	P04069 astacus flu
18	299.5	10.8	430	1	YHT2_YEAST	P38836 saccharomyc
19	277.5	10.0	424	1	CBPT_THEVU	P29068 thermoactin
20	277	10.0	451	1	CBPS_STRGR	P18143 streptomyce
21	252.5	9.1	434	1	CBPS_STRCP	P39041 streptomyce
22	118.5	4.3	707	1	ORC1_SCHPO	P54789 schizosacch
23	114.5	4.1	1380	1	CBPD_HUMAN	O75976 homo sapien
24	112.5	4.1	1377	1	CBPD_MOUSE	O89001 mus musculu
25	112	4.0	1968	1	RRPO_PVMR	P17965 potato viru
26	108.5	3.9	1255	1	PER2_HUMAN	O15055 homo sapien
27	103.5	3.7	722	1	CPXM_MOUSE	Q9z100 mus musculu
28	102	3.7	734	1	CPXM_HUMAN	Q96sm3 homo sapien
29	100	3.6	376	1	YQGT_BACSU	P54497 bacillus su
30	99.5	3.6	445	1	TPH1_CHICK	P70080 gallus gall
31	98.5	3.5	488	1	TY3H_ANGAN	O42091 anguilla an
32	98.5	3.5	1160	1	C8CA_BACTP	Q45706 bacillus th
33	97	3.5	663	1	TAZ1_SCHPO	P79005 schizosacch
34	95	3.4	444	1	TPH1_HUMAN	P17752 homo sapien
35	95	3.4	1793	1	YCF1_LOTJA	Q9bbn6 lotus japon
36	94.5	3.4	482	1	YPT1_CAEEL	P41879 caenorhabdi
37	94.5	3.4	744	1	YG13_YEAST	P53202 saccharomyc
38	94.5	3.4	2185	1	PYR1_DICDI	P20054 dictyosteli
39	94	3.4	435	1	YNJE_ECOLI	P78067 escherichia
40	93	3.3	853	1	PHS1_DICDI	Q00766 dictyosteli
41	92	3.3	525	1	PUR9_BUCAI	P57143 b bifunctio
42	92	3.3	835	1	VP3_ROTS1	P15736 simian 11 r
43	91.5	3.3	465	1	STHA_ECO57	Q8x727 escherichia
44	91.5	3.3	465	1	STHA_ECOLI	P27306 escherichia
45	91.5	3.3	982	1	ENV_SFV3L	P27399 simian foam

ALIGNMENTS

RESULT 1
CBPB_RAT
ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M23959; AAA40872.1; -.
DR EMBL; M23947; AAA40872.1; JOINED.
DR EMBL; M23950; AAA40872.1; JOINED.
DR EMBL; M23952; AAA40872.1; JOINED.
DR EMBL; M23953; AAA40872.1; JOINED.
DR EMBL; M23954; AAA40872.1; JOINED.
DR PIR; A32129; A32129.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -.

```
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    13       BY SIMILARITY.
FT   PROPEP      14   108       ACTIVATION PEPTIDE.
FT   CHAIN      109   415       CARBOXYPEPTIDASE B.
FT   METAL      174   174       ZINC (BY SIMILARITY).
FT   METAL      177   177       ZINC (BY SIMILARITY).
FT   METAL      302   302       ZINC (BY SIMILARITY).
FT   ACT_SITE   376   376       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   171   184       BY SIMILARITY.
FT   DISULFID   243   266       BY SIMILARITY.
FT   DISULFID   257   271       BY SIMILARITY.
SQ   SEQUENCE   415 AA;  47515 MW;  8EA06CCADE30B6F2 CRC64;

Alignment Scores:
Pred. No.:                6.32e-54      Length:          415
Score:                    695.50        Matches:         158
Percent Similarity:       52.74%        Conservative:    63
Best Local Similarity:    37.71%        Mismatches:      135
Query Match:              25.05%        Indels:          64
DB:                       1             Gaps:            8

US-09-980-881A-1 (1-1573) x CBPB_RAT (1-415)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||  |||          |||  |||  :::          |||:::|||
Db          1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGluGluHis--- 19

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
                 |||        :::|||                          |||
Db         20 ------PheAspGlyAsnArgValTyrArgValSerValHisGlyGluAspHisValAsn 37

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ::::::|||   |||      |||   ||||||      |||:::|||   ::|||
Db         38 LeuIleGlnGluLeuAlaAsnThrLysGluIleAspPheTrpLysProAspSerAlaThr 57

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              :::                  |||  |||  |||  |||  ||||||  :::|||
Db         58 GlnValLysProLeuThrThrValAspPheHisValLysAlaGluAspValAlaAspVal 77

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              :::      |||      :::  :::          ||||||:::::::::|||   :::
Db         78 GluAsnPheLeuGluGluAsnGluValHisTyrGluValLeuIleSerAsnValArgAsn 97

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ::::::   |||      :::  |||         |||||||||         |||
Db         98 AlaLeuGluSerGlnPheAspSerHisThr---------ArgAlaSerGlyHisSerTyr 114

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              :::|||:::             |||  :::||||||:::  :::         :::
Db        115 ThrLysTyrAsnLysTrpGluThrIleGluAlaTrpIleGlnGlnValAlaThrAspAsn 134

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||::::::|||:::      ||||||:::::||||||         :::||||||
Db        135 ProAspLeuValThrGlnSerValIleGlyThrThrPheGluGlyArgAsnMetTyrVal 154

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||:::   ||||||   :::      |||   ||||||:::|||||||||||||
Db        155 LeuLysIle---GlyLysThrArgProAsnLysProAlaIlePheIleAspCysGlyPhe 173

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------- 602
              ||||||||||||||||||||||||||||||||   ||||||:::
Db        174 HisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArg 193

Qy        602 ------------------------------------------------------------ 602

Db        194 ThrTyrAsnGlnGluIleHisMetLysGlnLeuLeuAspGluLeuAspPheTyrValLeu 213

Qy        603 ----------------GGCCAT----------------AATCGAATGTGGAGAAAG 626
                             |||:::                :::|||||||||||||||
Db        214 ProValValAsnIleAspGlyTyrValTyrThrTrpThrLysAspArgMetTrpArgLys 233

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
                 ||||||      |||   :::   |||:::|||          |||   ||||||
Db        234 ThrArgSerThrMetAlaGlySerSerCysLeuGlyValArgProAsnArgAsnPhe--- 252

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
                 :::      |||||||||   |||||||||   |||   ||||||||||||||||||||
Db        253 AsnAlaGlyTrpCysGluValGlyAlaSerArgSerProCysSerGluThrTyrCysGly 272

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
                  ||||||||||||   |||   ||||||:::|||   |||:::|||   |||:::
Db        273 ProAlaProGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeu 292

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
                 :::   ||||||||||||:::::::::|||||||||||||||||   :::::::::||||||
Db        293 SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrProTyr 312

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
                 ||||||                 :::::::::|||||||||:::   :::          |||
Db        313 SerTyrAspTyrLysLeuProGluAsnTyrGluGluLeuAsnAlaLeuValLysGlyAla 332

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
                 :::  :::      |||          |||:::||||||:::|||   |||:::
Db        333 AlaLysGluLeu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThr 351

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
                 |||:::|||   |||   ||||||   |||||||||   ||||||   ||||||||||||
Db        352 ThrIleTyrProAlaAlaGlyGlySerAspAspTrpSerTyrAspGlnGlyIleLysTyr 371

Qy       1047 TCGTTTACATC---------------------------------------------------- 1057
                 |||||||||
Db        372 SerPheThrPheGluLeuArgAspThrGlyPhePheGlyPheLeuLeuProGluSerGln 391

Qy       1058 ---AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTC 1111
                  :::   ||||||  |||          ||||||      ||||||   :::|||
Db        392 IleArgGlnThrCysGluGluThrMetLeuAlaValLysTyrIleAlaAsnTyrVal 410

RESULT 2
CBPC_MOUSE
ID   CBPC_MOUSE     STANDARD;      PRT;   417 AA.
AC   P15089;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90062123; PubMed=2584208;
RA   Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA   Serafin W.E.;
RT   "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT   novel member of the carboxypeptidase gene family.";
RL   J. Biol. Chem. 264:20094-20099(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; J05118; AAA37369.1; -.
DR   PIR; A34487; A34487.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   MGD; MGI:88479; Cpa3.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    15
FT   PROPEP      16   109       ACTIVATION PEPTIDE.
FT   CHAIN      110   417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL      176   176       ZINC (BY SIMILARITY).
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173   186       BY SIMILARITY.
FT   DISULFID   245   268       BY SIMILARITY.
SQ   SEQUENCE   417 AA;  48790 MW;  A2B300A068D1BA6D CRC64;

Alignment Scores:
Pred. No.:              2.54e-50      Length:         417
Score:                  655.00        Matches:        150
Percent Similarity:     51.67%        Conservative:   66
Best Local Similarity:  35.89%        Mismatches:     138
Query Match:            23.59%        Indels:         65
DB:                     1             Gaps:           9

US-09-980-881A-1 (1-1573) x CBPC_MOUSE (1-417)

Qy         33 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 92
              :::||||||:::         :::               |||         |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         93 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 152
                 :::|||                :::     :::         ||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        153 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 212
              |||   :::  |||:::     |||  |||          |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        213 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 272
                 |||  |||  |||:::     :::         :::::::::  |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        273 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 332
                   |||         :::|||:::   |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuHisAspLeuGlnGluGluIleGluLysGln 103

Qy        333 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 389
                  |||   ::::::   |||   |||         |||   :::|||:::
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
                :::::|||   ||||||   |||   :::   |||:::||||||:::|||:::
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
              ::::::|||   |||||||||:::   |||      ||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
              ||||||:::         :::   ||||||::::::||||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 608
              ||||||||||||||||||||   ||||||:::                     |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        609 AAT---------------------------------------------------------- 611
              |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        612 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 638
                                               ||||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        639 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 698
                  |||:::   ||||||||||||||||||||   ||||||          |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        699 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 758
                           ::::::      |||         |||   |||      ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        759 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 818
              ||||||   |||  |||||||||   :::|||:::|||   ::::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        819 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 878
              |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        879 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 938
                        :::|||::::::|||       ||||||       |||       |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        939 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 998
                    |||         |||||||||   :::|||       :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        999 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1054
                  |||         ||||||:::||||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1055 ----------------------------------------------ATCAAACCCACC 1066
                                                            ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuProGluSerArgIleLysProThr 397

Qy       1067 TGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
              |||:::|||         :::|||      ||||||   ::::::::::::|||
Db        398 CysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysAsn 415

RESULT 3
CBPB_HUMAN
ID   CBPB_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15086; O60834; Q96BQ8;
DT   01-APR-1990 (Rel. 14, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE   (PASP).
GN   CPB1 OR CPB OR PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC   TISSUE=Pancreas;
RX   MEDLINE=92129345; PubMed=1370825;
RA   Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA   French C.K.;
RT   "Isolation of a cDNA encoding a human serum marker for acute
RT   pancreatitis. Identification of pancreas-specific protein as
RT   pancreatic procarboxypeptidase B.";
RL   J. Biol. Chem. 267:2575-2581(1992).
RN   [2]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=98182241; PubMed=9524066;
RA   Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA   Aviles F.X.;
RT   "Comparative analysis of the sequences and three-dimensional models
RT   of human procarboxypeptidases A1, A2 and B.";
RL   Biol. Chem. 379:149-155(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 16-43.
RC   TISSUE=Pancreas;
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- TISSUE SPECIFICITY: Pancreas.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81057; AAA66973.1; -.
DR   EMBL; AJ224866; CAA12163.1; -.
DR   EMBL; BC015338; AAH15338.1; -.
DR   PDB; 1KWM; 28-AUG-02.
DR   MEROPS; M14.003; -.
DR   Genew; HGNC:2299; CPB1.
DR   MIM; 114852; -.
DR   GO; GO:0004180; F:carboxypeptidase activity; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1     15
FT   PROPEP      16    110       ACTIVATION PEPTIDE.
FT   CHAIN      111    417       CARBOXYPEPTIDASE B.
FT   METAL      176    176       ZINC (BY SIMILARITY).
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173    186       BY SIMILARITY.
FT   DISULFID   245    268       BY SIMILARITY.
FT   DISULFID   259    273       BY SIMILARITY.
FT   CONFLICT    16     16       H -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    17     17       H -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    37     37       H -> Q (IN REF. 4).
FT   CONFLICT   208    208       D -> N (IN REF. 2).
FT   CONFLICT   245    245       MISSING (IN REF. 1).
SQ   SEQUENCE   417 AA;  47367 MW;  EBBB98B27F5D5AF9 CRC64;

Alignment Scores:
Pred. No.:                4.23e-50        Length:          417
Score:                    652.50          Matches:         142
Percent Similarity:       52.13%          Conservative:    78
Best Local Similarity:    33.65%          Mismatches:      141
Query Match:              23.50%          Indels:          62
DB:                       1               Gaps:            8

US-09-980-881A-1 (1-1573) x CBPB_HUMAN (1-417)

Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
              :::   :::|||   |||||||||   :::   |||             |||
Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         84 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
                 |||:::    :::|||                          :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy        141 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 200
              ::::::   |||   :::|||   :::|||      |||:::|||   :::
Db         41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        201 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 260
              |||                 |||   |||   |||   |||   |||         |||:::
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy        261 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 320
                 |||      :::  :::      ||||||:::::::::::::   ::::::
Db         81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy        321 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 380
              ::::::   |||      :::               ||||||:::        ||||||
Db        101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
              :::|||:::            |||   :::|||   :::   :::         :::|||
Db        118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
                 ::::::::::::     ||||||::::::|||||         :::|||:::|||
Db        138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
              ||||||   ||||||           |||   ||||||::::::|||||||||   |||
Db        158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 602
              ||||||||||||||||||||||||||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy        602 ------------------------------------------------------------ 602

Db        197 TyrGlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPro 216

Qy        603 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629
                           |||:::                 :::|||   |||||||||
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy        630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
```

```
Db        237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
Db        256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
Db        316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
Db        336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
Db        355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

Qy       1050 TTTACATC-------------------------------------------------- 1057
Db        375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

Qy       1058 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGG 1117
Db        395 ArgAlaThrCysGluGluThrPheLeuAlaIleLysTyrValAlaSerTyrValLeuGlu 414

Qy       1118 AATGTT 1123
Db        415 HisLeu 416

RESULT 4
CBPC_HUMAN
ID   CBPC_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15088; Q96E94;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=90083291; PubMed=2594780;
RA   Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA   Austen K.F., Serafin W.E.;
RT   "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT   comparison of the protein with mouse mast cell carboxypeptidase A and
RT   rat pancreatic carboxypeptidases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mast cells;
RX   MEDLINE=92105393; PubMed=1729276;
RA   Reynolds D.S., Gurley D.S., Austen K.F.;
RT   "Cloning and characterization of the novel gene for mast cell
RT   carboxypeptidase A.";
RL   J. Clin. Invest. 89:273-282(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 110-417 FROM N.A.
RX   MEDLINE=92333165; PubMed=1629626;
RA   Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,
RA   Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT   "Human skin mast cell carboxypeptidase: functional characterization,
RT   cDNA cloning, and genealogy.";
RL   J. Invest. Dermatol. 99:138-145(1992).
RN   [5]
RP   SEQUENCE OF 110-137.
RX   MEDLINE=89214692; PubMed=2708524;
RA   Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT   "Human mast cell carboxypeptidase. Purification and
RT   characterization.";
RL   J. Clin. Invest. 83:1630-1636(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27717; AAA35652.1; -.
DR   EMBL; M73720; AAA59568.1; -.
DR   EMBL; M73716; AAA59568.1; JOINED.
DR   EMBL; M73717; AAA59568.1; JOINED.
DR   EMBL; M73718; AAA59568.1; JOINED.
DR   EMBL; M73719; AAA59568.1; JOINED.
DR   EMBL; BC012613; AAH12613.1; -.
DR   EMBL; S40234; AAB22578.2; ALT_SEQ.
DR   PIR; A43929; A43929.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   Genew; HGNC:2298; CPA3.
DR   MIM; 114851; -.
DR   GO; GO:0030141; C:secretory granule; NAS.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
```

```
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     15
FT   PROPEP       16    109       ACTIVATION PEPTIDE.
FT   CHAIN        110   417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL        176   176       ZINC (BY SIMILARITY).
FT   METAL        179   179       ZINC (BY SIMILARITY).
FT   METAL        304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE     378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     173   186       BY SIMILARITY.
FT   DISULFID     245   268       BY SIMILARITY.
FT   CONFLICT     63    63        N -> K (IN REF. 3).
SQ   SEQUENCE   417 AA;  48700 MW;  8CB90DB758117B24 CRC64;

Alignment Scores:
Pred. No.:                1.18e-49        Length:          417
Score:                    647.50          Matches:         143
Percent Similarity:       52.46%          Conservative:    81
Best Local Similarity:    33.49%          Mismatches:      132
Query Match:              23.32%          Indels:          72
DB:                       1               Gaps:            10

US-09-980-881A-1 (1-1573) x CBPC_HUMAN (1-417)

Qy          6 ATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGT 65
              :::   ||||||:::         :::||||||:::      |||:::
Db          5 LeuProValGlyLeuIleAlaThrThrLeuAlaIle---AlaProVal------------ 19

Qy         66 GAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 125
                              |||          :::|||           |||:::
Db         20 --------------ArgPheAspArgGluLysValPheArgValLysProGlnAspGlu 34

Qy        126 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 185
              :::|||     ::::::::::::|||    |||   |||:::     |||   |||
Db         35 LysGlnAlaAspIleIleLysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrPro 54

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                        :::               |||   |||   |||:::      :::
Db         55 GlyAlaThrHisHisValAlaAlaAsnMetMetValAspPheArgValSerGluLysGlu 74

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
                  :::::::::   |||:::   :::   :::         :::|||:::
Db         75 SerGlnAlaIleGlnSerAlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHis 94

Qy        306 GACGTGGAAGATCTTATTCAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCC 362
              |||::::::::::   |||::::::|||          :::   :::    |||
Db         95 AspLeuGlnGluGluIleGluLysGlnPheAspValLysGluAspIleProGlyArgHis 114

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              |||          |||   :::|||::::::      :::|||   :::|||   |||
Db        115 Ser---------TyrAlaLysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLys 131

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              :::   :::::::::|||:::|||::::::::::|||   |||||||||:::   |||
Db        132 MetMetAspLysTyrProGluMetValSerArgIleLysIleGlySerThrValGluAsp 151

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                 ||||||||||||||||||:::   |||::::::          :::   ||||||:::
Db        152 AsnProLeuTyrValLeuLysIle---GlyGluLysAsnGluArgArgLysAlaIlePhe 170

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
              :::|||||||||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        171 MetAspCysGlyIleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheVal 190

Qy        602 ------------------------------------------------------------ 602

Db        191 TyrGlnAlaThrLysThrTyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMet 210

Qy        603 ----------------------------------GGCCAT----------------AAT 611
                                               |||:::                |||
Db        211 AsnPheTyrIleLeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsn 230

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
              |||||||||||||||||||||||          |||:::   ||||||||||||||||||
```

```
Db        231 ArgMetTrpArgLysAsnArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeu 250

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
              |||   ||||||   :::     |||          ::::::      |||:::
Db        251 AsnArgAsnPhe---AsnAlaSerTrpAsnSerIleProAsnThrAsnAspProCysAla 269

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
              :::   |||   |||      ||||||||||||   |||   |||||||||   :::|||
Db        270 AspAsnTyrArgGlySerAlaProGluSerGluLysGluThrLysAlaValThrAsnPhe 289

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
              :::|||   ::::::|||:::||||||   ||||||:::   ||||||||||||||||
Db        290 IleArgSerHisLeuAsnGluIleLysValTyrIleThrPheHisSerTyrSerGlnMet 309

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
              ::::::||||||||||   ||||||               :::||||||:::|||:::
Db        310 LeuLeuProTyrGlyTyrThrSerLysLeuProProAsnHisGluAspLeuAlaLys 329

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
              ||||||                  :::      |||          |||||||||   :::
Db        330 ValAlaLysIleGlyThrAspValLeu---SerThrArgTyrGluThrArgTyrIleTyr 348

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
              |||            |||:::|||          |||          ||||||   ||||||
Db        349 GlyProIleGluSerThrIleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAsp 368

Qy       1032 TTGGGCATCAAATATTCGTTTAC---------------------------------- 1054
              ||||||||||||:::::::|||
Db        369 LeuGlyIleLysHisThrPheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeu 388

Qy       1055 ----------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATA 1099
                             ||||||||||||||||||||          ||||||      |||
Db        389 LeuProGluSerArgIleLysProThrCysArgGluThrMetLeuAlaValLysPheIle 408

Qy       1100 GCTTGGCATGTCATTAGGAAT 1120
              |||   ::::::::::::::
Db        409 AlaLysTyrIleLeuLysHis 415

RESULT 5
CBPB_CANFA
ID   CBPB_CANFA     STANDARD;      PRT;   416 AA.
AC   P55261;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE   membrane associated protein) (ZAP47).
GN   CPB1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Fukuoka S.-I.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D78348; BAA11366.1; -.
DR   HSSP; P09955; 1PBA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
```

```
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     15       BY SIMILARITY.
FT   PROPEP       16    109       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN       110    416       CARBOXYPEPTIDASE B.
FT   METAL       175    175       ZINC (BY SIMILARITY).
FT   METAL       178    178       ZINC (BY SIMILARITY).
FT   METAL       303    303       ZINC (BY SIMILARITY).
FT   ACT_SITE    377    377       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    244    267       BY SIMILARITY.
FT   DISULFID    258    272       BY SIMILARITY.
SQ   SEQUENCE   416 AA;  47595 MW;  DFE1992CD52F8AB4 CRC64;

Alignment Scores:
Pred. No.:                4.94e-49        Length:         416
Score:                    640.50          Matches:        142
Percent Similarity:       51.67%          Conservative:   75
Best Local Similarity:    33.81%          Mismatches:     136
Query Match:              23.06%          Indels:         68
DB:                       1               Gaps:           9

US-09-980-881A-1 (1-1573) x CBPB_CANFA (1-416)

Qy         39 GTCCTTGTACCCATTGTTCTC-----------TTCTGTGAGCAGCATGTCTTCGCGTTC 86
              :::||||||   :::   |||           :::     :::|||         |||
Db          5 IleLeuValThrLeuAlaLeuAlaSerAlaHisTyrSerGlyGluHis---------Phe 21

Qy         87 CAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAG 146
              :::     :::|||                 :::   :::|||
Db         22 GluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnLeuLeuHis 41

Qy        147 AATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTG 206
                |||   :::|||   :::|||      |||:::|||   :::         |||
Db         42 ThrLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLys 61

Qy        207 AAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCAT 266
                          |||   |||   |||   |||:::     |||:::
Db         62 ProHisSerThrAlaAspPheArgValLysAlaGluAspIleLeuThrValGluAspPhe 81

Qy        267 TTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAA 326
              |||     :::   :::          ||||||:::   ::::::     :::::::::
Db         82 LeuLysGlnAsnGluLeuHisTyrGluValLeuIleAsnAsnLeuArgLeuValLeuGlu 101

Qy        327 CAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTAT 386
                |||                          |||:::         ||||||:::|||
Db        102 GlyGlnPheGlyArgGlnVal---------ProAlaThrGlyHisSerTyrGluLysTyr 118

Qy        387 CACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATG 446
              :::          |||   :::|||   :::   :::|||      :::||||||:::
Db        119 AsnArgTrpGluThrIleGluAlaTrpThrGlnGlnValThrSerGluAsnProAspLeu 138

Qy        447 CTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTT 506
              :::::::::     ||||||::::::||||||         :::|||:::|||||||||
Db        139 IleSerArgArgSerIleGlyThrThrPheGluGlyArgThrIleTyrLeuLeuLysVal 158

Qy        507 TCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGA 566
                ||||||             |||   ||||||::::::|||||||||   |||||||||
Db        159 ---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArg 177

Qy        567 GAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
              ||||||||||||||||||||      ||||||:::
Db        178 GluTrpIleSerProAlaPheTrpGlnTrpPheValArgGlu***IleArgThrTyrGly 197

Qy        602 ------------------------------------------------------------ 602

Db        198 GlnGluIleHisMetThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValGly 217

Qy        603 ---------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCT 635
                      |||:::                 |||||||||||||||||   ||||||
Db        218 AsnIleAspGlyTyrValTyrThrTrpThrLysAsnArgMetTrpArgLysThrArgSer 237

Qy        636 TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACAC 695
                          :::|||:::|||||||||         ||||||
Db        238 ThrGlnValGlyThrAsnCysValGlyThrAspProThrArgAsnPhe---AspAlaGly 256

Qy        696 TGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCT 755
              ||||||:::   |||||||||   :::   |||   ||||||||||||||||
Db        257 TrpCysLysIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAla 276

Qy        756 GAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATT 815
              |||||||||   |||   ||||||:::|||:::|||:::|||   |||::::::   |||
Db        277 GluSerGluLysGluThrLysAlaLeuAlaAsnPheIleArgSerAsnLeuSerSerIle 296

Qy        816 AAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACA 875
              |||||||||:::::::::|||||||||||||||   :::::::::|||||||||||
Db        297 LysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrProTyrSerTyrAsp 316

Qy        876 CGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCT 935
                        :::::::::::   ||||||:::   :::|||         |||:::
Db        317 TyrLysLeuThrGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGlu 336

Qy        936 ATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATAC 995
              :::     |||      |||:::||||||:::|||   |||:::   |||:::|||
Db        337 Leu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyr 355

Qy        996 CTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1055
                |||   ||||||   |||||||||   ||||||   ||||||||||||||||||||
Db        356 ProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrSerPheThr 375

Qy       1056 TC-------------------------------------------------AAACCC 1063
                                                                   |||
Db        376 PheGluLeuArgAspLysGlyArgTyrGlyPheAlaLeuProGluSerGlnIleSerPro 395

Qy       1064 ACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123
              ||||||   |||         |||:::      :::|||   :::|||:::::::::::
Db        396 ThrCysGluGluThrLeuLeuAlaIleLysHisLeuAlaArgTyrValLeuGlnHisLeu 415

RESULT 6
CBPB_PIG
ID   CBPB_PIG       STANDARD;      PRT;   401 AA.
AC   P09955;
DT   01-MAR-1989 (Rel. 10, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2).
GN   CPB.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE OF 1-103.
RX   MEDLINE=91208150; PubMed=2018774;
RA   Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA   Aviles F.X.;
RT   "Analysis of the activation process of porcine procarboxypeptidase B
RT   and determination of the sequence of its activation segment.";
RL   Biochemistry 30:4082-4089(1991).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 1-38.
RX   MEDLINE=85279427; PubMed=4026847;
RA   Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT   "Sequential homologies between procarboxypeptidases A and B from
RT   porcine pancreas.";
RL   Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=91114690; PubMed=1989878;
RA   Coll M., Guasch A., Aviles F.X., Huber R.;
RT   "Three-dimensional structure of porcine procarboxypeptidase B: a
```

```
RT   structural basis of its inactivity.";
RL   EMBO J. 10:1-9(1991).
RN   [4]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX   MEDLINE=91027767; PubMed=2223783;
RA   Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT   "Sequence-specific 1H NMR assignments and determination of the
RT   secondary structure for the activation domain isolated from
RT   pancreatic procarboxypeptidase B.";
RL   Biochemistry 29:7515-7522(1990).
RN   [5]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=91114693; PubMed=1989879;
RA   Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
RT   "The NMR structure of the activation domain isolated from porcine
RT   procarboxypeptidase B.";
RL   EMBO J. 10:11-15(1991).
RN   [6]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=93044373; PubMed=1422143;
RA   Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA   Huber R., Wuethrich K.;
RT   "Comparison of the NMR solution structure with the X-ray crystal
RT   structure of the activation domain from procarboxypeptidase B.";
RL   J. Biomol. NMR 2:1-10(1992).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR   PDB; 1PBA; 31-OCT-93.
DR   PDB; 1NSA; 24-DEC-97.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW   3D-structure.
FT   PROPEP        1     95       ACTIVATION PEPTIDE.
FT   CHAIN         96    401      CARBOXYPEPTIDASE B.
FT   DISULFID      158   171
FT   DISULFID      230   253
FT   DISULFID      244   258
FT   METAL         161   161      ZINC.
FT   METAL         164   164      ZINC.
FT   METAL         289   289      ZINC.
FT   ACT_SITE      341   341
FT   ACT_SITE      363   363      NUCLEOPHILE.
FT   TURN            8     9
FT   STRAND         11    17
FT   HELIX          20    32
FT   STRAND         36    39
FT   HELIX          43    45
FT   STRAND         50    56
FT   HELIX          58    60
FT   HELIX          61    69
FT   TURN           70    72
FT   STRAND         74    79
FT   HELIX          82    87
FT   TURN           88    88
FT   TURN          101   102
FT   HELIX         107   120
FT   TURN          122   124
FT   STRAND        125   132
FT   TURN          134   135
FT   STRAND        138   144
FT   TURN          149   150
FT   STRAND        153   157
FT   TURN          162   163
FT   HELIX         165   181
FT   TURN          182   184
FT   HELIX         186   194
FT   STRAND        196   200
FT   HELIX         205   213
FT   TURN          214   214
FT   TURN          216   217
FT   STRAND        224   224
FT   TURN          226   227
FT   STRAND        231   231
FT   HELIX         235   237
FT   TURN          243   244
FT   STRAND        249   249
FT   TURN          252   253
FT   TURN          255   256
FT   STRAND        257   257
FT   TURN          262   263
FT   HELIX         266   277
FT   TURN          278   281
FT   STRAND        282   289
FT   STRAND        294   297
FT   TURN          307   308
FT   HELIX         309   327
FT   STRAND        332   335
FT   HELIX         336   339
FT   HELIX         347   353
FT   TURN          354   355
FT   STRAND        359   363
FT   TURN          371   372
FT   HELIX         376   378
FT   HELIX         379   398
FT   TURN          399   401
SQ   SEQUENCE   401 AA;  45713 MW;  53129AF159A26348 CRC64;

Alignment Scores:
Pred. No.:                4.67e-48        Length:          401
Score:                    629.50          Matches:         134
Percent Similarity:       54.85%          Conservative:    64
Best Local Similarity:    37.12%          Mismatches:      120
Query Match:              22.67%          Indels:          43
DB:                       1               Gaps:            6

US-09-980-881A-1 (1-1573) x CBPB_PIG (1-401)

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
              |||:::    :::|||                                  :::     |||
Db          7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerGluLeu 26

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
                  |||  :::|||  :::|||     |||:::|||  :::        |||
Db         27 HisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGlnIle 46

Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
                       |||  |||  |||  |||  |||:::     |||:::
Db         47 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 66

Qy        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
               |||      :::  :::         ||||||:::  ::::::     ::::::
Db         67 PheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerValLeu 86

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
              :::  |||     :::          |||  :::     |||||:::
Db         87 GluAlaGlnPheAspSerArgCys---------ArgThrThrGlyHisSerTyrGluLys 103

Qy        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
              |||::::::         |||  :::|||  |||  :::|||  ::::::||||||
Db        104 TyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValThrSerLysAsnProAsp 123

Qy        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
              ::::::::::::      ||||||:::::|||:::        :::|||:::||||||
Db        124 LeuIleSerArgSerAlaIleGlyThrThrPheAspGlyAspAsnIleTyrLeuLeuLys 143
```

```
Qy        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
              |||   ||||||      :::   |||   ||||||::::::|||||||||   ||||||
Db        144 Val---GlyLysProGlySerAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162

Qy        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
              ||||||||||||||   |||||||||   ||||||:::
Db        163 ArgGluTrpIleSerGlnAlaPheCysGlnTrpPheValArgAspAlaValArgThrTyr 182

Qy        602 ------------------------------------------------------------ 602

Db        183 GlyTyrGluAlaHisMetThrGluPheLeuAspAsnLeuAspPheTyrValLeuProVal 202

Qy        603 ------------GGCCAT------------------AATCGAATGTGGAGAAAGAACCGT 632
                          |||:::                  ||||||||||||||||||   |||
Db        203 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 222

Qy        633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
              |||      |||   :::   |||   |||||||||   |||   ||||||   :::
Db        223 SerThrAsnAlaGlySerSerCysThrGlyThrAspProAsnArgAsnPhe---AsnAla 241

Qy        693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
                  ||||||      |||||||||   :::   |||:::|||||||||||||||
Db        242 GlyTrpCysThrValGlyAlaSerValAsnProCysAsnGluThrTyrCysGlySerAla 261

Qy        753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
              ||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        262 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSerSer 281

Qy        813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
              |||||||||||:::::::::|||||||||||||||   |||:::::|||||||||||||
Db        282 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 301

Qy        873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
                      ::::::      ||||||:::   :::|||      ||||||:::
Db        302 AspTyrLysLeuProGluAsnAspAlaGluLeuAsnSerLeuAlaLysGlyAlaValLys 321

Qy        933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
              :::     :::       |||   |||::::::|||   ||||||   |||:::
Db        322 GluLeu---AlaSerLeuTyrGlyThrSerTyrSerTyrGlyProGlySerThrThrIle 340

Qy        993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
              |||   |||   ||||||   |||||||||   |||:::   ||||||||||||||||||
Db        341 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAsnGlnGlyIleLysTyrSerPhe 360

Qy       1053 ACA 1055
              |||
Db        361 Thr 361
```

RESULT 7
CBPB_BOVIN
ID CBPB_BOVIN STANDARD; PRT; 306 AA.
AC P00732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN [2]
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX MEDLINE=74260705; PubMed=4833744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cystines and identification of the active site
RT arginine.";
RL J. Biol. Chem. 249:3756-3764(1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Herriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190(1976).
RN [4]
RP ACTIVE SITE.
RX MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 244:5246-5253(1969).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 247:7864-7869(1972).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A93797; CPBOB.
DR PDB; 1CPB; 30-SEP-83.
DR MEROPS; M14.003; -.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 66 66 ZINC.
FT METAL 69 69 ZINC.
FT METAL 194 194 ZINC.
FT ACT_SITE 246 246
FT ACT_SITE 268 268 NUCLEOPHILE.
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;

Alignment Scores:

Pred. No.:	1.61e-44	Length:	306
Score:	589.50	Matches:	121
Percent Similarity:	56.58%	Conservative:	51
Best Local Similarity:	39.80%	Mismatches:	75
Query Match:	21.23%	Indels:	58
DB:	1	Gaps:	6

US-09-980-881A-1 (1-1573) x CBPB_BOVIN (1-306)

```
Qy        375 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 434
              ||||||:::|||::::::          |||   :::|||   |||   :::
Db          6 TyrGluLysTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25

Qy        435 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 494
              :::||||||::::::::::::       ||||||::::::|||            :::|||
Db         26 AsnProAspLeuIleSerArgSerAlaIleGlyThrThrPheLeuGlyAsnThrIleTyr 45

Qy        495 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 554
              :::|||||||||   ||||||      :::   |||   |||:::::::::|||||||||
Db         46 LeuLeuLysVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64

Qy        555 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 602
                  ||||||||||||||||||||||||||||||||   ||||||:::
Db         65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84

Qy        602 ------------------------------------------------------------ 602
```

```
Db         85 ArgThrTyrGlyArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104

Qy        603 --------------------GGCCAT-----------------AATCGAATGTGGAGA 623
                              |||:::                 ||||||||||||||
Db        105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrThrAsnArgMetTrpArg 124

Qy        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683
              |||   ||||||      |||   :::   |||   ||||||||||||||   ||||||
Db        125 LysThrArgSerThrArgAlaGlySerSerCysThrGlyThrAspLeuAsnArgAsnPhe 144

Qy        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
                  ||||||      |||||||||::::::   ||||||||||||||||||
Db        145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163

Qy        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803
              |||         |||||||||   |||   ||||||||||||   |||:::|||   :::
Db        164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgAsnHis 183

Qy        804 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 863
              ::::::   ||||||||||||:::::::::|||||||||||||||   :::::::::|||
Db        184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203

Qy        864 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 923
              |||||||||                |||::::::   ||||||:::   :::|||
Db        204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223

Qy        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983
              ||||||:::   :::      :::          |||   |||::::::|||   |||:::
Db        224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242

Qy        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1043
              |||:::|||  |||   ||||||   |||||||||   ||||||   |||||||||
Db        243 ThrThrIleTyrProAlaSerGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLys 262

Qy       1044 TATTCGTTTACATC---------------------------------------------- 1057
              ||||||||||||
Db        263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282

Qy       1058 ------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTC 1111
                    :::|||||||||   |||         |||:::      :::      :::|||
Db        283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIleLysTyrValThrSerTyrVal 302

Qy       1112 ATTAGGAATGTT 1123
              :::   ::::::
Db        303 LeuGluHisLeu 306

RESULT 8
CBPC_RAT
ID   CBPC_RAT       STANDARD;      PRT;   309 AA.
AC   P21961;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE   A3).
GN   CPA3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91105153; PubMed=1988052;
RA   Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA   Neurath H.;
RT   "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT   enzyme activity within mast cell granules.";
RL   Biochemistry 30:648-655(1991).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A38395; A38395.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   METAL        68     68       ZINC (BY SIMILARITY).
FT   METAL        71     71       ZINC (BY SIMILARITY).
FT   METAL       196    196       ZINC (BY SIMILARITY).
FT   ACT_SITE    248    248       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    270    270       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     65     78       BY SIMILARITY.
FT   DISULFID    137    160       BY SIMILARITY.
FT   VARIANT       1      1       MISSING (IN THE MAJOR FORM).
SQ   SEQUENCE   309 AA;  35786 MW;  20330FABC3EE83EF CRC64;

Alignment Scores:
Pred. No.:              1.46e-42        Length:         309
Score:                  567.50          Matches:        120
Percent Similarity:     54.79%          Conservative:   46
Best Local Similarity:  39.60%          Mismatches:     80
Query Match:            20.44%          Indels:         58
DB:                     1               Gaps:           6

US-09-980-881A-1 (1-1573) x CBPC_RAT (1-309)

Qy        375 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 434
              |||   :::|||:::      |||:::|||   ||||||   |||   :::   |||:::
Db          8 TyrAlaLysTyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLys 27

Qy        435 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 494
              ||||||:::|||:::::::::|||   ||||||||:::   |||       |||||||||
Db         28 HisProGluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyr 47

Qy        495 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 554
              |||||||||:::   |||::::::        :::   ||||||::::::|||||||||
Db         48 ValLeuLysIle---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66

Qy        555 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 602
              |||||||||||||||||:::|||||||||||||||   ||||||:::
Db         67 IleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86

Qy        603 ---------GGCCATAAT---------------------------------------- 611
                       |||   |||
Db         87 LysSerTyrGlyLysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106

Qy        612 ----------------------------------------------------CGAATGTGGAGA 623
                                                               ||||||||||||
Db        107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126

Qy        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683
              ||||||||||||         ::::::   ||||||||||||||||||||   ||||||
Db        127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146

Qy        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
                        |||                 :::   :::   |||         |||
Db        147 ---AspValSerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArg 165

Qy        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803
              |||      ||||||||||||   |||   |||||||||   :::|||:::|||   :::
Db        166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185

Qy        804 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 863
              :::|||   |||||||||||||||:::   ||||||||||||||   ::::::||||||
Db        186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205

Qy        864 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 923
              |||   |||||                :::|||::::::|||      ||||||
Db        206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIle 225
```

```
Qy        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983
              |||          :::       :::          ||||||||   :::|||      :::
Db        226 AlaThrAspValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAla 244

Qy        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1043
                 |||:::|||          |||         ||||||   ||||||||||||||||||
Db        245 SerThrIleTyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLys 264

Qy       1044 TATTCGTTTAC---------------------------------------------------- 1054
                 ::::::|||
Db        265 HisThrPheAlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSer 284

Qy       1055 ---ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTC 1111
                  ||||||||||||||:::|||         :::|||       ||||||   ::::::
Db        285 ArgIleLysProThrCysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIle 304

Qy       1112 ATTAGGAAT 1120
                 ::::::::: .
Db        305 LeuLysHis 307

RESULT 9
CBP2_RAT
ID   CBP2_RAT       STANDARD;      PRT;   417 AA.
AC   P19222;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=89034323; PubMed=3182871;
RA   Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA   Graf M., Rutter W.J.;
RT   "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT   cloning, and evolutionary implications on substrate specificity in
RT   the carboxypeptidase gene family.";
RL   J. Biol. Chem. 263:17828-17836(1988).
RN   [2]
RP   SEQUENCE OF 131-143 FROM N.A.
RX   MEDLINE=95386501; PubMed=7657630;
RA   Normant E., Gros C., Schwartz J.C.;
RT   "Carboxypeptidase A isoforms produced by distinct genes or
RT   alternative splicing in brain and other extrapancreatic tissues.";
RL   J. Biol. Chem. 270:20543-20549(1995).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=92105124; PubMed=1761558;
RA   Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT   "Structural evolution of an enzyme specificity. The structure of rat
RT   carboxypeptidase A2 at 1.9-A resolution.";
RL   J. Biol. Chem. 266:24606-24612(1991).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23721; AAA40956.1; -.
DR   EMBL; M23714; AAA40956.1; JOINED.
DR   EMBL; M23715; AAA40956.1; JOINED.
DR   EMBL; M23716; AAA40956.1; JOINED.
DR   EMBL; M23717; AAA40956.1; JOINED.
DR   EMBL; M23718; AAA40956.1; JOINED.
DR   EMBL; M23719; AAA40956.1; JOINED.
DR   EMBL; M23720; AAA40956.1; JOINED.
DR   EMBL; S79837; -; NOT_ANNOTATED_CDS.
DR   PIR; A32128; A32128.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.002; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16       POTENTIAL.
FT   PROPEP      17    112       ACTIVATION PEPTIDE.
FT   CHAIN      113    417       CARBOXYPEPTIDASE A2.
FT   METAL      177    177       ZINC.
FT   METAL      180    180       ZINC.
FT   METAL      304    304       ZINC.
FT   ACT_SITE   378    378       NUCLEOPHILE.
FT   DISULFID   246    269       BY SIMILARITY.
FT   DISULFID   318    352       PROBABLE.
SQ   SEQUENCE   417 AA;  46912 MW;  BEDCC41A830F2D45 CRC64;

Alignment Scores:
Pred. No.:                3.01e-41        Length:          417
Score:                    553.00          Matches:         137
Percent Similarity:       50.37%          Conservative:    69
Best Local Similarity:    33.50%          Mismatches:      137
Query Match:              19.91%          Indels:          66
DB:                       1               Gaps:            13

US-09-980-881A-1 (1-1573) x CBP2_RAT (1-417)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||:::|||      |||         :::         ::::::|||::::::
Db          1 MetArgLeuThrLeuLeuLeuAlaAlaLeuLeuGlyTyrIleTyrCysGlnGlu------ 18

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
                 |||          |||||||||       :::|||              |||::::::
Db         19 ---ThrPheValGlyAspGlnValLeuGluIleIleProSerHisGluGluGlnIleArg 37

Qy        138 GTTCTACAGAATCTTACTACAACA------TATGAGATTGTTCTCTGGCAGCCGGTAACA 191
                 |||      |||                  |||:::      |||:::      |||
Db         38 ThrLeuLeuGlnLeuGluAlaGluGluHisLeuGluLeuAspPheTrpLysSerProThr 57

Qy        192 GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGAC 251
                           :::       :::   ||||||       |||       :::   :::
Db         58 ------------IleProGlyGluThrValHisValArgValProPheAlaSerIleGln 73

Qy        252 AATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTG 311
                 ||||||      |||          ||||||       |||:::::::::   ||||||
Db         74 AlaValLysValPheLeuGluSerGlnGlyIleAspTyrSerIleMetIleGluAspVal 93

Qy        312 GAAGATCTTATTCAACAGCAGATTTCCAAC---------GACACAGTCAGCCCCCGAGCC 362
                 :::   |||:::   |||:::                 :::                |||
Db         94 GlnValLeuLeuAspGlnGluArgGluGluMetLeuPheAsnGlnGlnArgGluArgGly 113

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
                           :::|||   ||||||:::|||   |||||||||      ::::::
Db        114 GlyAsnPheAsnPheGluAlaTyrHisThrLeuGluGluIleTyrGlnGluMetAspAsn 133

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
                 :::         :::|||   :::::::::|||:::::::::|||||||||||||||
Db        134 LeuValAlaGluAsnProGlyLeuValSerLysValAsnLeuGlySerSerPheGluAsn 153

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACAGCCAAAAATGCCATA 539
                 |||:::   |||||||||   |||   |||   :::         |||   ||||||
```

```
Db        154 ArgProMetAsnValLeuLysPheSerThrGlyGlyAsp---------LysProAlaIle 170

Qy        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGG--- 596
              |||:::|||   |||||||||||||||||||:::::   |||      ||||||
Db        171 TrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAlaThrAlaLeuTrpThr 190

Qy        596 ------------------------------------------------------------ 596

Db        191 AlaAsnLysIleAlaSerAspTyrGlyThrAspProAlaIleThrSerLeuLeuAsnThr 210

Qy        597 ---------TTCATA--------------------GGCCAT----------------- 608
                      |||:::                   |||:::
Db        211 LeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrValPheSerGlnThrThr 230

Qy        609 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 668
              ||||||||||||||||||   ||||||      :::   :::   |||:::|||   |||
Db        231 AsnArgMetTrpArgLysThrArgSerLysArgSerGlySerGlyCysValGlyValAsp 250

Qy        669 CTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 728
                 |||   |||:::         ::::::          ||||||||||||||   |||
Db        251 ProAsnArgAsnTrp---AspAlaAsnPheGlyGlyProGlyAlaSerSerSerProCys 269

Qy        729 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 788
              |||::::::|||   |||      |||   ||||||   |||||||||::::::
Db        270 SerAspSerTyrHisGlyProLysProAsnSerGluValGluValLysSerIleValAsp 289

Qy        789 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 848
              |||::::::   :::      ::::::||||||:::|||::::::|||||||||||||||
Db        290 PheIleLysSerHis---GlyLysValLysAlaPheIleThrLeuHisSerTyrSerGln 308

Qy        849 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 908
                 ::::::|||||||||   |||      :::|||      |||      ||||||
Db        309 LeuLeuMetPheProTyrGlyTyrLysCysThrLysProAspAspPheAsnGluLeuAsp 328

Qy        909 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACA 968
                 ||||||   :::|||   :::|||:::   |||              |||   |||
Db        329 GluValAlaGlnLysAlaAlaGlnAlaLeu---LysArgLeuHisGlyThrSerTyrLys 347

Qy        969 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1028
                 |||                 :::|||   |||   ||||||      ||||||   |||
Db        348 ValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySerIleAspTrpAlaTyr 367

Qy       1029 GATTTGGGCATCAAATATTCGTTTACATCAAAC-------------------------- 1061
              |||||||||||||||||||||||||
Db        368 AspLeuGlyIleLysTyrSerPheAlaPheGluLeuArgAspThrAlaPheTyrGlyPhe 387

Qy       1062 ---CCACCTGTAGAGAAGCTTTTGCCG 1085
                  |||   :::::::::||||||
Db        388 LeuLeuProAlaLysGlnIleLeuPro 396
```

```
RESULT 10
CBPA_BOVIN
ID   CBPA_BOVIN     STANDARD;      PRT;   419 AA.
AC   P00730;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A precursor (EC 3.4.17.1).
GN   CPA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91151335; PubMed=1998496;
RA   le Hueerou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
RT   "Cloning and nucleotide sequence of a bovine pancreatic
RT   preprocarboxypeptidase A cDNA.";
RL   Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=96096549; PubMed=8522204;
RA   Goo J.H., Kim K.H., Choi K.Y.;
RT   "Cloning, sequencing and expression of the gene encoding a major
RT   allotypic preprocarboxypeptidase A from bovine pancreas.";
RL   Gene 165:333-334(1995).
RN   [3]
RP   SEQUENCE OF 111-417.
RX   MEDLINE=71155183; PubMed=5102489;
RA   Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT   chymotryptic peptides of the cyanogen bromide fragment F-I.";
RL   Biochemistry 10:938-950(1971).
RN   [4]
RP   REVISIONS TO 138 AND 141.
RX   MEDLINE=72138789; PubMed=5143102;
RA   Petra P.H., Hermodson M.A., Walsh K.A., Neurath H.;
RT   "Characterization of bovine carboxypeptidase A (Allan).";
RL   Biochemistry 10:4023-4025(1971).
RN   [5]
RP   SEQUENCE OF 17-120.
RX   MEDLINE=89150306; PubMed=3147705;
RA   Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neurath H.;
RT   "The amino acid sequence of the activation peptide of bovine pro-
RT   carboxypeptidase A.";
RL   Biochimie 70:1137-1142(1988).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
RX   MEDLINE=83294519; PubMed=6887246;
RA   Rees D.C., Lewis M., Lipscomb W.N.;
RT   "Refined crystal structure of carboxypeptidase A at 1.54-A
RT   resolution.";
RL   J. Mol. Biol. 168:367-387(1983).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX   MEDLINE=96003618; PubMed=7556081;
RA   Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT   "The three-dimensional structure of the native ternary complex of
RT   bovine pancreatic procarboxypeptidase A with proproteinase E and
RT   chymotrypsinogen C.";
RL   EMBO J. 14:4387-4394(1995).
RN   [8]
RP   VARIANT ALLELIC.
RX   MEDLINE=69283620; PubMed=5817619;
RA   Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Identification of the amino acid replacements characterizing the
RT   allotypic forms of bovine carboxypeptidase A.";
RL   Biochemistry 8:2762-2768(1969).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC       COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC       PROPROTEINASE E.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M61851; AAA30426.1; -.
DR   EMBL; M61851; AAA30427.1; -.
DR   EMBL; Z33906; CAA83955.1; -.
DR   PIR; JN0126; CPBOA.
DR   PDB; 3CPA; 15-JAN-87.
DR   PDB; 4CPA; 22-OCT-84.
DR   PDB; 5CPA; 15-JAN-87.
DR   PDB; 6CPA; 15-OCT-91.
```

```
DR   PDB; 7CPA; 31-JAN-94.
DR   PDB; 8CPA; 31-JAN-94.
DR   PDB; 1CBX; 31-JAN-94.
DR   PDB; 1CPS; 15-OCT-94.
DR   PDB; 2CTB; 31-JAN-94.
DR   PDB; 2CTC; 31-JAN-94.
DR   PDB; 1ARL; 01-AUG-96.
DR   PDB; 1ARM; 17-AUG-96.
DR   PDB; 1BAV; 01-APR-97.
DR   PDB; 1YME; 12-FEB-97.
DR   PDB; 1CPX; 05-AUG-98.
DR   PDB; 1PYT; 27-JAN-97.
DR   PDB; 1ELL; 26-JUN-02.
DR   PDB; 1ELM; 26-JUN-02.
DR   PDB; 1PS7; 06-SEP-00.
DR   PDB; 1HDQ; 15-NOV-01.
DR   PDB; 1HDU; 15-NOV-01.
DR   PDB; 1HEE; 23-NOV-01.
DR   PDB; 1M4L; 04-MAR-03.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure; Polymorphism.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A.
FT   METAL      179   179       ZINC.
FT   METAL      182   182       ZINC.
FT   METAL      306   306       ZINC.
FT   ACT_SITE   358   358       PROTON DONOR.
FT   ACT_SITE   380   380       NUCLEOPHILE.
FT   DISULFID   248   271
FT   VARIANT    289   289       I -> V (IN ALLELIC VARIANT).
FT   VARIANT    338   338       E -> A (IN ALLELIC VARIANT).
FT   VARIANT    415   415       L -> V (IN ALLELIC VARIANT).
FT   CONFLICT    95    95       S -> L (IN REF. 5).
FT   CONFLICT   199   199       D -> N (IN REF. 3).
FT   CONFLICT   203   203       D -> N (IN REF. 3).
FT   CONFLICT   224   224       D -> N (IN REF. 3).
FT   CONFLICT   232   232       Q -> E (IN REF. 3).
FT   CONFLICT   295   295       D -> N (IN REF. 3).
FT   TURN       114   116
FT   TURN       119   120
FT   HELIX      125   138
FT   TURN       140   142
FT   STRAND     143   150
FT   TURN       152   153
FT   STRAND     156   162
FT   STRAND     171   176
FT   TURN       180   181
FT   HELIX      183   199
FT   TURN       200   202
FT   HELIX      204   212
FT   STRAND     214   218
FT   HELIX      223   231
FT   TURN       232   232
FT   TURN       234   235
FT   STRAND     242   242
FT   TURN       244   245
FT   STRAND     249   249
FT   TURN       253   254
FT   TURN       261   262
FT   STRAND     267   267
FT   TURN       270   271
FT   TURN       273   274
FT   STRAND     275   275
FT   TURN       280   281
FT   HELIX      284   296
FT   STRAND     299   307
FT   STRAND     311   314
FT   TURN       324   325
FT   HELIX      326   344
FT   STRAND     349   352
FT   HELIX      353   356
FT   TURN       357   357
FT   HELIX      364   370
FT   TURN       371   372
FT   STRAND     375   381
FT   TURN       388   389
FT   HELIX      393   395
FT   HELIX      396   416
SQ   SEQUENCE   419 AA;  47082 MW;  21B86407B3BFC452 CRC64;

Alignment Scores:
Pred. No.:                9.79e-40        Length:          419
Score:                    536.00          Matches:         141
Percent Similarity:       47.80%          Conservative:    65
Best Local Similarity:    32.71%          Mismatches:      145
Query Match:              19.30%          Indels:          81
DB:                       1               Gaps:            12

US-09-980-881A-1 (1-1573) x CBPA_BOVIN (1-419)

Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
              |||      |||:::||||||:::          |||      |||
Db          4 LeuLeuIleLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp---------- 19

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
              |||         |||||||||                          :::||||||   :::
Db         20 PheValGlyHisGlnValLeuArgIleThrAlaAlaAspGluAlaGluValGlnThrVal 39

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGG------------CAGCCG 185
              :::   |||               :::   |||      |||            ||||||
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProGlyGlnPro 59

Qy        186 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 245
                                                :::        |||
Db         60 ------------------------GlySerProIleAspValArgValProPheProSer 71

Qy        246 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 305
              :::      ||||||      |||         ||||||        :::::::::
Db         72 LeuGlnAlaValLysValPheLeuGluAlaHisGlyIleArgTyrArgIleMetIleGlu 91

Qy        306 GACGTGGAAGATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCC 356
              ||||||:::   |||:::          |||:::|||:::
Db         92 AspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaSerGlnSerArgAla 111

Qy        357 CGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 413
              |||::::::   ::::::   |||      ||||||:::|||:::|||||||||   :::
Db        112 ArgSerThrAsnThrPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrAspPhe 131

Qy        414 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 473
              ::::::   :::         ||||||   :::::::::|||:::   ||||||   |||
Db        132 MetAspLeuLeuValAlaGluHisProGlnLeuValSerLysLeuGlnIleGlyArgSer 151

Qy        474 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 533
              :::|||      |||:::||||||||||||   |||           :::   :::
Db        152 TyrGluGlyArgProIleTyrValLeuLysPheSer---ThrGlyGlySerAsnArgPro 170

Qy        534 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 593
              ||||||||||||||   |||||||||:::||||||||||||:::   |||      :::
Db        171 AlaIleTrpIleAspLeuGlyIleHisSerArgGluTrpIleThrGlnAlaThrGlyVal 190

Qy        594 TGGTTCATA---------------------------------------------------- 602
              ||||||
Db        191 TrpPheAlaLysLysPheThrGluAspTyrGlyGlnAspProSerPheThrAlaIleLeu 210

Qy        602 ---------------------------------------------------------- 602
```

```
Db        211 AspSerMetAspIlePheLeuGluIleValThrAsnProAspGlyPheAlaPheThrHis 230

Qy        603 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 662
              ||||||:::|||||||||   ||||||      :::::::::   .|||:::|||
Db        231 SerGlnAsnArgLeuTrpArgLysThrArgSerValThrSerSerSerLeuCysValGly 250

Qy        663 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 722
              |||   |||   |||:::           :::   :::   |||||||||||||||
Db        251 ValAspAlaAsnArgAsnTrp---AspAlaGlyPheGlyLysAlaGlyAlaSerSerSer 269

Qy        723 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 782
              |||||||||||||||   |||   |||      ||||||   ||||||||::::::
Db        270 ProCysSerGluThrTyrHisGlyLysTyrAlaAsnSerGluValGluValLysSerIle 289

Qy        783 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 842
              |||::::::   :::   |||      ||||||::::::|||:::|||||||||
Db        290 ValAspPheValLysAspHisGlyAsn---PheLysAlaPheLeuSerIleHisSerTyr 308

Qy        843 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 902
              ||||||   :::::::::||||||   ||||||               |||      |||
Db        309 SerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrGlnSerIleProAspLysThrGlu 328

Qy        903 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 962
              |||:::   ||||||      ||||||   |||:::   |||:::      |||
Db        329 LeuAsnGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyrGlyThrSer 347

Qy        963 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1022
              |||   :::|||           |||:::|||   |||   ||||||      ||||||
Db        348 TyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSerGlyGlySerIleAspTrp 367

Qy       1023 ATCTATGATTTGGGCATCAAATATTCGTTTACATCAAA---------------------- 1060
              |||:::   ||||||||||||||||||||
Db        368 SerTyrAsnGlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyArgTyr 387

Qy       1061 ------------------------------CCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1090
                                           ||||||   :::|||   :::      |||
Db        388 GlyPheLeuLeuProAlaSerGlnIleIleProThrAlaGlnGluThrTrpLeuGlyVal 407

Qy       1091 TCTAAAATAGCTTGGCATGTCATTAGGAATGTT 1123
              |||      |||   :::   |||:::
Db        408 LeuThrIleMetGluHisThrLeuAsnAsnLeu 418
```

```
RESULT 11
CBP1_RAT
ID   CBP1_RAT       STANDARD;      PRT;   419 AA.
AC   P00731;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82105986; PubMed=6275388;
RA   Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA   Pictet R.L., Valenzuela P., Rutter W.J.;
RT   "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT   characterization of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89034324; PubMed=3182872;
RA   Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT   "Structural characterization of the rat carboxypeptidase A1 and B
RT   genes. Comparative analysis of the rat carboxypeptidase gene
RT   family.";
RL   J. Biol. Chem. 263:17837-17845(1988).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01232; CAA24542.1; -.
DR   EMBL; J00713; AAA40893.1; -.
DR   EMBL; M23990; AAA40955.1; -.
DR   EMBL; M23960; AAA40955.1; JOINED.
DR   EMBL; M23985; AAA40955.1; JOINED.
DR   EMBL; M23986; AAA40955.1; JOINED.
DR   EMBL; M23987; AAA40955.1; JOINED.
DR   EMBL; M23988; AAA40955.1; JOINED.
DR   EMBL; M23989; AAA40955.1; JOINED.
DR   PIR; A00911; CPRTA.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A1.
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      182   182       ZINC (BY SIMILARITY).
FT   METAL      306   306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358   358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380   380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248   271       BY SIMILARITY.
FT   CONFLICT   196   196       I -> V (IN REF. 1).
FT   CONFLICT   261   263       FGM -> LGK (IN REF. 1).
FT   CONFLICT   347   347       K -> E (IN REF. 1).
SQ   SEQUENCE   419 AA;  47197 MW;  BB002D1CB99B7491 CRC64;
```

```
Alignment Scores:
Pred. No.:              3.35e-39        Length:         419
Score:                  530.00          Matches:        140
Percent Similarity:     48.61%          Conservative:   70
Best Local Similarity:  32.41%          Mismatches:     154
Query Match:            19.09%          Indels:         68
DB:                     1               Gaps:           12

US-09-980-881A-1 (1-1573) x CBP1_RAT (1-419)

Qy         21 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 80
              :::|||      |||::::::|||:::     |||      |||
Db          3 ArgLeuLeuIleLeuSerLeuLeuLeuGluAlaValCysGlyAsnGluAsn--------- 19

Qy         81 GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
                 |||         |||||||||                              |||||||||
Db         20 ---PheValGlyHisGlnValLeuArgIleSerAlaAlaAspGluAlaGlnValGlnLys 38

Qy        141 CTACAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAGCCGGTAACAGCT 194
              ::::::   |||             :::   |||      |||:::
Db         39 ValLysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg 58

Qy        195 GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT 254
                    |||          :::|||   |||                  :::   :::
Db         59 AlaGlyIleProIleAspValArgValProPhe------------ProSerIleGlnSer 74

Qy        255 GTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAA 314
              |||||||||  |||         ||||||         :::::::::   ||||||:::
```

```
Db         75 ValLysAlaPheLeuGluTyrHisGlyIleSerTyrGluIleMetIleGluAspValGln 94

Qy        315 GATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCC 365
                 |||:::         :::||||||:::|||                    ::::::
Db         95 LeuLeuLeuAspGluGluLysGlnGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr 114

Qy        366 GCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
                 |||:::   |||      ||||||:::|||:::|||||||||   :::::::::
Db        115 AspSerPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrGluPheMetAspLeu 134

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
                 :::         ||||||   :::::::::||||||   ||||||::::::||||||
Db        135 LeuValAlaGluHisProGlnLeuValSerLysIleGlnIleGlyAsnThrPheGluGly 154

Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
                 |||::::::|||||||||   |||              |||   :::   |||||||||
Db        155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTrp 173

Qy        543 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 602
                 ||||||   |||||||||:::|||||||||::::::   |||      :::||||||
Db        174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193

Qy        603 GGCCAT---------------------------------------------------- 608

Db        194 LysLysIleThrLysAspTyrGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213

Qy        609 ----------------------------------------------------------AAT 611
                                                                         |||
Db        214 AspIlePheLeuGluIleValThrAsnProAspGlyPheAlaTyrThrHisLysThrAsn 233

Qy        612 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 671
                 ||||||||||||||   ||||||          :::   |||:::|||   |||
Db        234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuCysValGlyValAspPro 253

Qy        672 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 731
                 |||   |||:::           :::         ||||||||||||:::   ||||||
Db        254 AsnArgAsnTrp---AspAlaGlyPheGlyMetAlaGlyAlaSerSerAsnProCysSer 272

Qy        732 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 791
                 |||||||||   |||   :::|||   ||||||   |||||||||::::::      |||
Db        273 GluThrTyrArgGlyLysPheProAsnSerGluValGluValLysSerIleValAspPhe 292

Qy        792 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 851
                 :::      :::   |||   |||||||||:::||||||:::||||||||||||||
Db        293 ValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSerTyrSerGlnLeu 311

Qy        852 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 911
                 :::::::::||||||   ||||||          :::   |||       ||||||
Db        312 LeuLeuTyrProTyrGlyTyrThrSerGluProAlaProAspGlnAlaGluLeuAspGln 331

Qy        912 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 971
                 :::|||      ||||||   |||:::      :::         |||::::::   :::
Db        332 LeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThrLysPheLysTyr 350

Qy        972 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1031
                 |||         :::|||:::|||   |||   |||         ||||||   |||
Db        351 GlySerIleIleAspThrIleTyrGlnAlaSerGlySerThrIleAspTrpThrTyrSer 370

Qy       1032 TTGGGCATCAAATATTCGTTTACATCAAAC---------------------------- 1061
                 ||||||||||||||||||||
Db        371 GlnGlyIleLysTyrSerPheThrPheGluLeuArgAspThrGlyLeuArgGlyPheLeu 390

Qy       1062 CCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATG 1121
                 |||      :::::::::|||   :::   :::                |||   :::
Db        391 LeuProAlaSerGlnIleIleProThrAlaGluGlu---------ThrTrpLeuAlaLeu 407

Qy       1122 TTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
                      |||      :::      ||||||
Db        408 LeuThrIleMetAspHisThrValLysHisProTyr 419

RESULT 12
CPB2_HUMAN
ID   CPB2_HUMAN     STANDARD;      PRT;   417 AA.
AC   P48052; Q96A12; Q96QN3;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC   TISSUE=Pancreas;
RX   MEDLINE=95204457; PubMed=7896805;
RA   Catasus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
RA   Billeter M.;
RT   "The sequence and conformation of human pancreatic
RT   procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT   three-dimensional model.";
RL   J. Biol. Chem. 270:6651-6657(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98111000; PubMed=9450539;
RA   Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,
RA   Aviles F.X.;
RT   "Characterisation and preliminary X-ray diffraction analysis of human
RT   pancreatic procarboxypeptidase A2.";
RL   FEBS Lett. 420:7-10(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98046021; PubMed=9384570;
RA   Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT   "The three-dimensional structure of human procarboxypeptidase A2.
RT   Deciphering the basis of the inhibition, activation and intrinsic
RT   activity of the zymogen.";
RL   EMBO J. 16:6906-6913(1997).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U19977; AAA74425.1; -.
DR   EMBL; BC007009; AAH07009.1; -.
DR   EMBL; BC014571; AAH14571.1; -.
DR   EMBL; BC015140; AAH15140.1; -.
DR   PIR; A56171; A56171.
DR   PDB; 1AYE; 13-JAN-99.
DR   PDB; 1DTD; 12-JUL-00.
DR   PDB; 1O6X; 30-JAN-03.
DR   MEROPS; M14.002; -.
DR   Genew; HGNC:2297; CPA2.
DR   MIM; 600688; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0007039; P:vacuolar protein catabolism; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   112       ACTIVATION PEPTIDE.
FT   CHAIN      113   417       CARBOXYPEPTIDASE A2.
FT   METAL      177   177       ZINC.
FT   METAL      180   180       ZINC.
FT   METAL      304   304       ZINC.
FT   ACT_SITE   378   378       NUCLEOPHILE.
FT   DISULFID   246   269
FT   DISULFID   318   352
FT   CONFLICT    80    80       E -> G (IN REF. 2; AAH07009).
FT   CONFLICT   302   302       T -> I (IN REF. 1).
SQ   SEQUENCE   417 AA;  46828 MW;  3D9C073A98700A83 CRC64;

Alignment Scores:
Pred. No.:              1.64e-37        Length:         417
Score:                  511.00          Matches:        128
Percent Similarity:     48.73%          Conservative:   64
Best Local Similarity:  32.49%          Mismatches:     136
Query Match:            18.40%          Indels:         66
DB:                     1               Gaps:           11

US-09-980-881A-1 (1-1573) x CPB2_HUMAN (1-417)

Qy          72 CATGTCTTC---------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACC 122
               |||::::::         |||      |||||||||     :::|||
Db          13 HisIleTyrCysLeuGluThrPheValGlyAspGlnValLeuGluIleValProSerAsn 32

Qy         123 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC------ 176
                  |||::::::   |||      |||               :::   |||
Db          33 GluGluGlnIleLysAsnLeuLeuGlnLeuGluAlaGlnGluHisLeuGlnLeuAspPhe 52

Qy         177 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 236
               |||:::      |||                  :::      |||      |||
Db          53 TrpLysSerProThr-----------ThrProGlyGluThrAlaHisValArgValPro 68

Qy         237 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 296
                  :::|||     ||||||      |||          ||||||      |||:::
Db          69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTyrSerIle 88

Qy         297 TTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCC 356
               ::::::   ||||||:::   |||:::   ::::::                  :::
Db          89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108

Qy         357 CGAGCCTCCGCATCG---------TACTATGAACAGTATCACTCACTAAATGAAATCTAT 407
               |||          |||            :::      ||||||:::|||   ||||||
Db         109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyrHisThrLeuGluGluIleSer 128

Qy         408 TCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGA 467
                     ::::::   :::      ||||||   :::::::::|||::::::||||||
Db         129 GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148

Qy         468 TCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACA 524
               ||||||||||||      |||:::   |||||||||   |||   |||   :::
Db         149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp------ 166

Qy         525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
                  |||   |||||||||:::|||   |||||||||||||||||||||:::::   |||
Db         167 ---LysProAlaIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185

Qy         585 TTCTGCTTGTGGTTCATAGGC---------------------------------------- 605
                     ||||||
Db         186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205

Qy         605 ---------------------------------------------------------- 605

Db         206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225

Qy         606 ------------CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 653
                          ||||||||||||||||||   ||||||      :::   :::
Db         226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu 245

Qy         654 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 713
               |||:::|||   |||   |||   |||:::            :::         ||||||
Db         246 CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264

Qy         714 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 773
               ||||||:::   ||||||::::::|||   |||              ||||||   ||||||
Db         265 SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal 284

Qy         774 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 833
               |||::::::      |||::::::   :::      ::::::||||||:::|||::::::
Db         285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleThrLeu 303

Qy         834 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 893
               |||||||||||||||   ::::::|||||||||   |||      :::|||      |||
Db         304 HisSerTyrSerGlnLeuLeuMetPheProTyrGlyTyrLysCysThrLysLeuAspAsp 323

Qy         894 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 953
                  :::|||||||||   ||||||   :::|||   :::::::::   ::::::
Db         324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342

Qy         954 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 1013
               |||:::|||      |||                  :::|||   |||   ||||||
Db         343 GlyThrLysTyrLysValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer 362

Qy        1014 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC------------ 1061
               ||||||   ||||||   ||||||||||||||||||
Db         363 IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr 382

Qy        1062 ------------------CCACCTGTAGAGAAGCTTTTGCCG 1085
                                   |||      ::::::||||||
Db         383 GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro 396

RESULT 13
CBP1_HUMAN
ID   CBP1_HUMAN     STANDARD;      PRT;   419 AA.
AC   P15085; Q9BS67;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1 OR CPA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=93038569; PubMed=1417781;
RA   Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA   Wicker-Planquart C., Puigserver A.;
RT   "cDNA cloning and sequence analysis of human pancreatic
```

```
RT   procarboxypeptidase A1.";
RL   Biochem. J. 287:299-303(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 17-42.
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67318; CAA47732.1; -.
DR   EMBL; BC005279; AAH05279.1; -.
DR   PIR; S29127; S29127.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   Genew; HGNC:2296; CPA1.
DR   MIM; 114850; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A1.
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      182   182       ZINC (BY SIMILARITY).
FT   METAL      306   306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358   358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380   380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248   271       BY SIMILARITY.
FT   CONFLICT   276   276       H -> R (IN REF. 2).
SQ   SEQUENCE   419 AA;  47140 MW;  439FAFFFAEE958B1 CRC64;

Alignment Scores:
Pred. No.:              1.34e-35        Length:         419
Score:                  489.50          Matches:        131
Percent Similarity:     47.63%          Conservative:   60
Best Local Similarity:  32.67%          Mismatches:     127
Query Match:            17.63%          Indels:         83
DB:                     1               Gaps:           13

US-09-980-881A-1 (1-1573) x CBP1_HUMAN (1-419)

Qy        24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
             |||      |||:::||||||:::      |||         |||
Db         4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluAsp---------- 19

Qy        84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
             |||         |||||||||                                |||||||||   :::
Db        20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValGlnLysVal 39

Qy       144 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAG-------------- 182
             :::   |||          :::   |||      |||:::
Db        40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59

Qy       183 ------CCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 236
                 |||:::                          :::|||   |||
Db        60 GlySerProIleAspVal--------------------ArgValProPhe--------- 69

Qy       237 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 296
                   :::     ||||||      |||         |||||
Db        70 ---ProSerIleGlnAlaValLysIlePheLeuGluSerHisGlyIleSerTyrGluThr 88

Qy       297 TTGCTGGCAGACGTGGAAGATCTTATT---------CAACAGCAGATT------------ 335
             ::::::   ||||||:::   |||:::          |||:::|||:::
Db        89 MetIleGluAspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaPheArg 108

Qy       336 ------------TCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
                        |||   ||||||   ::                        |||
Db       109 SerArgAlaArgSerThrAspThrPheAsn--------------------TyrAlaThr 121

Qy       384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
             |||||:::|||   |||||||||   :::::::::   :::          :::|||
Db       122 TyrHisThrLeuGluGluIleTyrAspPheLeuAspLeuLeuValAlaGluAsnProHis 141

Qy       444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
             :::::::::||||||   ||||||:::::::::|||      |||:::||||||||||||
Db       142 LeuValSerLysIleGlnIleGlyAsnThrTyrGluGlyArgProIleTyrValLeuLys 161

Qy       504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
                |||            :::   :::   |||||||||||||||   |||||||||:::
Db       162 PheSer---ThrGlyGlySerLysArgProAlaIleTrpIleAspThrGlyIleHisSer 180

Qy       564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------- 602
             |||||||||::::::   |||      :::||||||
Db       181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyr 200

Qy       602 ------------------------------------------------------------ 602

Db       201 GlyGlnAspAlaAlaPheThrAlaIleLeuAspThrLeuAspIlePheLeuGluIleVal 220

Qy       603 ----------------------------GGCCATAATCGAATGTGGAGAAAGAACCGT 632
                                         ||||||||||||||||||   |||
Db       221 ThrAsnProAspGlyPheAlaPheThrHisSerThrAsnArgMetTrpArgLysThrArg 240

Qy       633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
             |||      |||   :::   |||||||||   |||   |||   |||:::
Db       241 SerHisThrAlaGlySerLeuCysIleGlyValAspProAsnArgAsnTrp---AspAla 259

Qy       693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
                :::         |||||||||||:::   ||||||||||||||   |||   :::
Db       260 GlyPheGlyLeuSerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyLysPhe 279
```

```
Qy        753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
                  ||||||   |||||||||::::::      |||::::::   :::   |||
Db        280 AlaAsnSerGluValGluValLysSerIleValAspPheValLysAspHisGlyAsn--- 298

Qy        813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
              ||||||||:::||||||::||||||||||||||   :::::::::||||||   |||
Db        299 IleLysAlaPheIleSerIleHisSerTyrSerGlnLeuLeuMetTyrProTyrGlyTyr 318

Qy        873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
                        |||   :::||||||      ::::::      ||||||
Db        319 LysThrGluProValProAspGlnAspGluLeuAspGlnLeuSerLysAlaAlaValThr 338

Qy        933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
              |||:::      :::          |||::::::   :::|||          :::   :::
Db        339 AlaLeu---AlaSerLeuTyrGlyThrLysPheAsnTyrGlySerIleIleLysAlaIle 357

Qy        993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
              |||   |||   |||          ||||||   |||       ||||||||||||||||
Db        358 TyrGlnAlaSerGlySerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377

Qy       1053 ACA 1055
              |||
Db        378 Thr 378
```

```
RESULT 14
CBP4_HUMAN
ID   CBP4_HUMAN     STANDARD;      PRT;   421 AA.
AC   Q9UI42;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN   CPA4 OR CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99310537; PubMed=10383164;
RA   Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT   "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT   deacetylase inhibitors during differentiation of prostate epithelial
RT   cancer cells.";
RL   Cancer Res. 59:2981-2988(1999).
RN   [2]
RP   IMPRINTING.
RX   MEDLINE=22483629; PubMed=12552318;
RA   Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA   Yoshiura K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA   Niikawa N., Kishino T.;
RT   "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT   imprinting domain.";
RL   Hum. Genet. 112:220-226(2003).
CC   -!- FUNCTION: Could be involved in the histone hyperacetylation
CC       pathway.
CC   -!- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,
CC       heart, intestine, kidney, liver and lung. Except for fetal brain
CC       that shows no imprinting, expression was found preferentially from
CC       the maternal allele.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095719; AAF23230.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.017; -.
DR   Genew; HGNC:15740; CPA4.
DR   MIM; 607635; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; NAS.
DR   GO; GO:0016573; P:histone acetylation; NAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1    16       POTENTIAL.
FT   PROPEP       17   113       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN       114   421       CARBOXYPEPTIDASE A4.
FT   METAL       181   181       ZINC (BY SIMILARITY).
FT   METAL       184   184       ZINC (BY SIMILARITY).
FT   METAL       308   308       ZINC (BY SIMILARITY).
FT   ACT_SITE    382   382       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    250   273       BY SIMILARITY.
SQ   SEQUENCE   421 AA;  47379 MW;  91883D25B705BC6A CRC64;
```

```
Alignment Scores:
Pred. No.:              5.08e-35        Length:         421
Score:                  483.00          Matches:        119
Percent Similarity:     48.92%          Conservative:   63
Best Local Similarity:  31.99%          Mismatches:     130
Query Match:            17.39%          Indels:         60
DB:                     1               Gaps:           10

US-09-980-881A-1 (1-1573) x CBP4_HUMAN (1-421)

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
              |||       |||||||||           |||       ::::::       |||
Db         21 PhePheGlyAspGlnValLeuArgIleAsnValArgAsnGlyAspGluIleSerLysLeu 40

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAGCCGGTAACAGCT--- 194
                 |||      :::      :::   |||      |||:::      ::::::
Db         41 SerGlnLeuValAsnSerAsnAsnLeuLysLeuAsnPheTrpLysSerProSerSerPhe 60

Qy        195 ----------GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCT 242
                       |||::::::|||           :::       |||
Db         61 AsnArgProValAspValLeuValProSerValSerLeuGlnAlaPhe---------- 76

Qy        243 GATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTG 302
                          |||:::   |||       |||:::      :::|||   :::
Db         77 --------------LysSerPheLeuArgSerGlnGlyLeuGluTyrAlaValThrIle 91

Qy        303 GCAGACGTGGAAGATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGC 353
                 |||::::::   |||:::               ::::::   ::::::
Db         92 GluAspLeuGlnAlaLeuLeuAspAsnGluAspAspGluMetGlnHisAsnGluGlyGln 111

Qy        354 CCCCGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCT 410
                 |||:::|||   ::::::   |||      ||||||||||||      ||||||
Db        112 GluArgSerSerAsnAsnPheAsnTyrGlyAlaTyrHisSerLeuGluAlaIleTyrHis 131

Qy        411 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 470
                 ::::::   |||          ||||||:::      ::::::   ||||||
Db        132 GluMetAspAsnIleAlaAlaAspProProAspLeuAlaArgArgValLysIleGlyHis 151

Qy        471 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAA 530
                 |||||||||      |||:::||||||||||||   |||       :::       :::
Db        152 SerPheGluAsnArgProMetTyrValLeuLysPheSerThrGlyLysGlyValArgArg 171

Qy        531 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGC 590
                 |||:::|||::::::   |||||||||:::|||||||||||||||   |||
Db        172 ProAlaValTrpLeuAsnAlaGlyIleHisSerArgGluTrpIleSerGlnAlaThrAla 191

Qy        591 TTGTGG---------------------------------------------------------- 596
                  :::|||
Db        192 IleTrpThrAlaArgLysIleValSerAspTyrGlnArgAspProAlaIleThrSerIle 211
```

```
Qy        597 ------------------TTCATA----------------------------------- 602
                              |||:::
Db        212 LeuGluLysMetAspIlePheLeuLeuProValAlaAsnProAspGlyTyrValTyrThr 231

Qy        603 ---GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 659
                  ||||||:::|||||||||   ||||||            :::   ||||||
Db        232 GlnThrGlnAsnArgLeuTrpArgLysThrArgSerArgAsnProGlySerSerCysIle 251

Qy        660 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 719
              |||   |||   |||   |||:::  :::      :::      :::|||||||||
Db        252 GlyAlaAspProAsnArgAsnTrp---AsnAlaSerPheAlaGlyLysGlyAlaSerAsp 270

Qy        720 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 779
              :::   |||||||||   |||   |||   :::      ||||||   |||||||||:::
Db        271 AsnProCysSerGluValTyrHisGlyProHisAlaAsnSerGluValGluValLysSer 290

Qy        780 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 839
              |||      |||:::::::::::   |||      |||   :::|||   :::||||||
Db        291 ValValAspPheIleGlnLysHisGlyAsn---PheLysGlyPheIleAspLeuHisSer 309

Qy        840 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 899
              ||||||||   :::::::::||||||   |||:::     |||:::   |||   |||
Db        310 TyrSerGlnLeuLeuMetTyrProTyrGlyTyrSerValLysLysAlaProAspAlaGlu 329

Qy        900 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959
              ||||||      ||||||      |||   :::|||:::         |||      |||
Db        330 GluLeuAspLysValAlaArgLeuAlaAlaLysAlaLeuAlaSerValSer---GlyThr 348

Qy        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019
                |||      |||            |||:::|||   |||   |||         |||
Db        349 GluTyrGlnValGlyProThrCysThrThrValTyrProAlaSerGlySerSerIleAsp 368

Qy       1020 TGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1055
              |||   ||||||   |||||||||::::::||||||
Db        369 TrpAlaTyrAspAsnGlyIleLysPheAlaPheThr 380

RESULT 15
CBPZ_SIMVI
ID   CBPZ_SIMVI     STANDARD;      PRT;   304 AA.
AC   P42788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS   Simulium vittatum (Black fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC   Simuliidae; Simulium.
OX   NCBI_TaxID=7192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Gut;
RX   MEDLINE=94093864; PubMed=8269093;
RA   Ramos A., Mahowald A., Jacobs-Lorena M.;
RT   "Gut-specific genes from the black fly Simulium vittatum encoding
RT   trypsin-like and carboxypeptidase-like proteins.";
RL   Insect Mol. Biol. 1:149-163(1993).
CC   -!- FUNCTION: Involved in the digestion of the blood meal.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Gut-specific.
CC   -!- INDUCTION: By blood meal.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08481; AAA18531.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.UPA; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   NON_TER       1      1
FT   METAL        58     58       ZINC (BY SIMILARITY).
FT   METAL        61     61       ZINC (BY SIMILARITY).
FT   METAL       184    184       ZINC (BY SIMILARITY).
FT   ACT_SITE    236    236       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    259    259       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    125    148       BY SIMILARITY.
SQ   SEQUENCE   304 AA;  34849 MW;  2E6E3FF8A6AA9144 CRC64;

Alignment Scores:
Pred. No.:              1.84e-28        Length:         304
Score:                  409.00          Matches:        98
Percent Similarity:     49.68%          Conservative:   55
Best Local Similarity:  31.82%          Mismatches:     95
Query Match:            14.73%          Indels:         60
DB:                     1               Gaps:           10

US-09-980-881A-1 (1-1573) x CBPZ_SIMVI (1-304)

Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
              |||||||||:::|||   ||||||||||||||:::::   :::   :::   ||||||
Db          1 GlnTyrHisThrLeuProGluIleTyrSerTrpLeuAspArgLeuValGlnGluHisPro 20

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
              :::   :::      :::      |||   |||:::|||         :::      :::
Db         21 GluHisValGluProValValGlyGlyLysSerTyrGluGlyArgGluIleArgGlyVal 40

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
              |||||||||   |||:::                  :::   ::::::      ||||||
Db         41 LysValSerTyrLysLys------GlyAsnProValValMetValGluSerAsnIleHis 58

Qy        561 GCCAGAGAATGGATCTCTCCTGCT---------TTCTGCTTG------------------ 593
              ||||||||||||||:::   |||            :::   |||
Db         59 AlaArgGluTrpIleThrAlaAlaThrThrThrTyrLeuLeuAsnGluLeuLeuThrSer 78

Qy        594 ------------------------------------TGGTTCATA------------ 602
                                                 |||:::|||
Db         79 LysAsnSerThrIleArgGluMetAlaGluAsnTyrAspTrpTyrIleProValThr 98

Qy        603 ---------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCT 635
                      |||:::                 :::||||||||||||||   ||||||
Db         99 AsnProAspGlyTyrValTyrThrHisThrThrAspArgMetTrpArgLysThrArgSer 118

Qy        636 TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACAC 695
                    ::::::   |||   |||||||||   |||   |||:::   :::   |||
Db        119 ProAsnProAspSerLeuCysAlaGlyThrAspProAsnArgAsnTrp---AsnPheHis 137

Qy        696 TGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCT 755
              |||   |||:::|||   ||||||      |||:::|||||||||   |||
Db        138 TrpMetGluGlnGlyThrSerSerArgProCysThrGluThrTyrGlyGlyLysLysAla 157

Qy        756 GAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATT 815
                ||||||   |||   :::::   :::   ||||||:::                ||||||
Db        158 PheSerGluValGluThrArgSerPheSerAspPheLeuLysThrLeuLysGlyGlnIle 177

Qy        816 AAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACA 875
              |||   |||::::::   |||||||||||||||   ::::::||||||||||   :::|||
Db        178 LysValTyrLeuAlaPheHisSerTyrSerGlnLeuLeuLeuProTyrGlyHisThr 197

Qy        876 CGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCT 935
                  :::   :::|||::::::|||      :::         |||   |||:::
Db        198 CysGlnHisThrTyrAsnHisAspAspLeuGlnAlaIleGlyAspAlaAlaAlaArgSer 217

Qy        936 ATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATAC 995
              :::   :::            |||   ||||||   |||:::      :::   :::|||
```

```
Db        218 LeuAlaGln---ArgTyrGlyThrAspTyrThrValGlyAsnIleTyrAspAlaIleTyr 236

Qy        996 CTAGCTCCTGGAGGTGGGGACGATTGGATCTAT---------GATTTGGGCATCAAATAT 1046
                 |||   ||||||      ||||||   |||          |||:::   |||   |||
Db        237 ProAlaSerGlyGlySerMetAspTrpAlaTyrAspThrLeuAspIleProIleAlaTyr 256

Qy       1047 TCGTTTACATCAAACCCA-------------------------CCTGTAGAGAAGCTT 1079
              ::::::          |||                       |||      ::::::
Db        257 ThrTyrGluLeuArgProArgAspGlyTrpAsnGlyPheGlnLeuProAlaAsnGlnIle 276

Qy       1080 TTGCCG-------------------------CTGTCTCTAAAATAGCTTGGCATGTCA 1112
              :::|||                        :::::::::
Db        277 IleProThrGlyGluGluThrValAspSerValValThrIleLeuLysGluSerArgArg 296

Qy       1113 TTAGGAATGTTTAATGCCCCTGAT 1136
              ||||||   ||||||    |||
Db        297 LeuGlyTyrPheAsnThrSerAsp 304


Search completed: January  1, 2004, 18:53:49
Job time : 40.6016 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:10:36 ; Search time 78.8575 Seconds
(without alignments)
10294.941 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 2777
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135748_24954/app_query.fasta_1.2958
-DB=SPTREMBL_23 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_190_@runat_31122003_135748_24954 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1885	67.9	423	4	Q15114	Q15114 homo sapien
2	1884	67.8	360	4	Q9P2Y6	Q9p2y6 homo sapien
3	1880	67.7	423	4	Q96IY4	Q96iy4 homo sapien
4	1592.5	57.3	422	11	Q9JHH6	Q9jhh6 mus musculu
5	1586.5	57.1	422	11	Q9QZF0	Q9qzf0 mus musculu
6	1568.5	56.5	422	11	Q9EQV9	Q9eqv9 rattus norv
7	1029	37.1	198	4	Q9NTI8	Q9nti8 homo sapien
8	714	25.7	416	13	Q9PUF2	Q9puf2 bothrops ja
9	703.5	25.3	408	13	Q8AXN3	Q8axn3 paralichthy
10	646	23.3	416	6	Q9XSP3	Q9xsp3 sus scrofa
11	642.5	23.1	412	11	P97597	P97597 rattus norv
12	633.5	22.8	437	4	Q8N4T0	Q8n4t0 homo sapien
13	625	22.5	434	13	Q8AVS2	Q8avs2 xenopus lae
14	581	20.9	419	13	Q8UUK1	Q8uuk1 gallus gall
15	570	20.5	374	4	Q8IVL8	Q8ivl8 homo sapien
16	556	20.0	436	11	Q8R4H4	Q8r4h4 mus musculu
17	534	19.2	419	13	Q8AXN4	Q8axn4 paralichthy
18	520	18.7	436	4	Q8WXQ8	Q8wxq8 homo sapien
19	516.5	18.6	419	6	Q9TV85	Q9tv85 sus scrofa
20	507.5	18.3	420	11	Q8BMK6	Q8bmk6 mus musculu
21	504	18.1	262	11	Q8BVD0	Q8bvd0 mus musculu
22	490	17.6	452	13	Q8UW45	Q8uw45 fugu rubrip
23	484	17.4	247	4	Q96KZ9	Q96kz9 homo sapien
24	461	16.6	428	5	Q8IGC4	Q8igc4 drosophila
25	460.5	16.6	419	13	Q8AXN5	Q8axn5 paralichthy
26	460	16.6	424	5	Q9VL87	Q9vl87 drosophila
27	459	16.5	1132	5	Q9W475	Q9w475 drosophila
28	451.5	16.3	422	5	Q9VLZ2	Q9vlz2 drosophila
29	440.5	15.9	226	4	Q8TDE8	Q8tde8 homo sapien
30	436	15.7	584	5	Q9TZH1	Q9tzh1 caenorhabdi
31	434.5	15.6	351	5	Q8T0Q1	Q8t0q1 drosophila
32	434	15.6	351	4	Q8NA08	Q8na08 homo sapien
33	418	15.1	445	5	Q9W478	Q9w478 drosophila
34	416.5	15.0	488	5	Q9BL88	Q9bl88 caenorhabdi
35	414.5	14.9	416	5	Q9VRZ3	Q9vrz3 drosophila
36	407.5	14.7	419	5	Q9VLZ1	Q9vlz1 drosophila
37	403	14.5	430	5	Q8SZP6	Q8szp6 drosophila
38	402	14.5	424	5	Q9VS66	Q9vs66 drosophila
39	402	14.5	430	5	Q9VL86	Q9vl86 drosophila
40	400.5	14.4	540	5	Q9TZH2	Q9tzh2 caenorhabdi
41	399.5	14.4	427	5	Q9U9K2	Q9u9k2 aedes aegyp
42	397	14.3	323	5	Q19121	Q19121 caenorhabdi
43	396	14.3	453	5	Q9V342	Q9v342 drosophila
44	386.5	13.9	423	5	O61532	O61532 drosophila
45	382.5	13.8	581	5	Q23318	Q23318 caenorhabdi

ALIGNMENTS

RESULT 1
Q15114
ID Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.009; -.
DR Genew; HGNC:2300; CPB2.

```
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
FT   CHAIN        23    423       PLASMA CARBOXYPEPTIDASE B.
SQ   SEQUENCE   423 AA;  48442 MW;  9B383272F6EE79F4 CRC64;

Alignment Scores:
Pred. No.:              5.21e-181      Length:        423
Score:                  1885.00        Matches:       365
Percent Similarity:     86.52%         Conservative:  1
Best Local Similarity:  86.29%         Mismatches:    3
Query Match:            67.88%         Indels:        55
DB:                     4              Gaps:          2

US-09-980-881A-1 (1-1573) x Q15114 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 -------------------------------------------AATCGAATGTGGAGAAAG 626
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT----------------------------------------------------TAC 1054
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 2
Q9P2Y6
ID   Q9P2Y6      PRELIMINARY;      PRT;   360 AA.
AC   Q9P2Y6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B-like protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Matsumoto A.;
RT   "Isolation, molecular cloning, and partial characterization of a novel
RT   carboxypeptidase B from human plasma.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92042093; PubMed=1939207;
RA   Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT   "Isolation, molecular cloning, and partial characterization of a novel
RT   carboxypeptidase B from human plasma.";
RL   J. Biol. Chem. 266:21833-21838(1991).
DR   EMBL; AB011969; BAA90475.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   360 AA;  40935 MW;  BF670B2F7437C1CB CRC64;

Alignment Scores:
Pred. No.:              6.32e-181      Length:        360
Score:                  1884.00        Matches:       354
```

```
Percent Similarity:     98.89%          Conservative:  2
Best Local Similarity:  98.33%          Mismatches:    4
Query Match:            67.84%          Indels:        0
DB:                     4               Gaps:          0

US-09-980-881A-1 (1-1573) x Q9P2Y6 (1-360)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnThrGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy        678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977
              ||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1038 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleLysTyrSerPheThrSerAsnProProValGluLysLeuProLeuSerLeuLys 360

RESULT 3
Q96IY4
ID   Q96IY4      PRELIMINARY;      PRT;   423 AA.
AC   Q96IY4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B2 (Plasma).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007057; AAH07057.1; -.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   423 AA;  48412 MW;  9B383E03F6EE7CF5 CRC64;

Alignment Scores:
Pred. No.:              1.66e-180       Length:         423
Score:                  1880.00         Matches:        364
Percent Similarity:     86.29%          Conservative:   1
Best Local Similarity:  86.05%          Mismatches:     4
Query Match:            67.70%          Indels:         55
DB:                     4               Gaps:           2

US-09-980-881A-1 (1-1573) x Q96IY4 (1-423)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        608 ------------------------------------------------------------ 608

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        609 -----------------------------------------AATCGAATGTGGAGAAAG 626
                                                      ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1047 TCGTT-------------------------------------------------------TAC 1054
              |||                                                       |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLysIleAlaTrpHisValIle 420

Qy       1115 AGGAATGTT 1123
              |||||||||
Db        421 ArgAsnVal 423

RESULT 4
Q9JHH6
ID   Q9JHH6      PRELIMINARY;      PRT;   422 AA.
AC   Q9JHH6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE   (1110032P04Rik protein).
GN   CPB2 OR TAFI OR 1110032P04RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20341711; PubMed=10878383;
RA   Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA   Campbell W., Okada H.;
RT   "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT   whereas carboxypeptidase N is not.";
RL   J. Immunol. 165:1053-1058(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20201996; PubMed=10739389;
RA   Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA   van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT   "Characterization of mouse thrombin-activatable fibrinolysis
RT   inhibitor.";
RL   Thromb. Haemost. 83:297-303(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AB021968; BAB03402.1; -.
DR   EMBL; AF164524; AAF62385.1; -.
DR   EMBL; AK004045; BAB23141.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48870 MW;  99113755669D55CB CRC64;

Alignment Scores:
Pred. No.:              1.63e-151      Length:          422
Score:                  1592.50        Matches:         309
Percent Similarity:     79.15%         Conservative:    25
Best Local Similarity:  73.22%         Mismatches:      33
Query Match:            57.35%         Indels:          56
DB:                     11             Gaps:            3

US-09-980-881A-1 (1-1573) x Q9JHH6 (1-422)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||||||||       |||   :::||||||   |||:::|||:::   |||||||||
Db          1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db         20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              :::|||||||||||||||||||||||||||:::|||||||||||||||||||||||:::
```

```
Db        40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy       198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
Db        60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79

Qy       258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
Db        80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
Db       100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
Db       120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy       438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
Db       140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159

Qy       498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
Db       160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy       558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
Db       180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy       608 ---------------------------------------------------------- 608

Db       200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy       609 -------------------------------------------AATCGAATGTGGAGAAAG 626
Db       220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
Db       240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
Db       260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 279

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
Db       280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
Db       300 AspHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db       320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
Db       340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
Db       360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy      1047 TCGTT----------------------------------------------------TAC 1054
Db       380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399

Qy      1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
Db       400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLysIleValTrpHisValIle 419
```

```
Qy      1115 AGGAAT 1120
Db       420 ArgAsn 421
```

RESULT 5
Q9QZF0
ID Q9QZF0 PRELIMINARY; PRT; 422 AA.
AC Q9QZF0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase U.
GN CPB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAF1J;
RA He Y.C., Broze G.;
RT "Isolation and characterization of mouse liver carboxypeptidase B
RT gene.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF186188; AAF00528.1; -.
DR HSSP; P00730; 5CPA.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48832 MW; B53FFFB09943954E CRC64;

Alignment Scores:

Pred. No.:	6.57e-151	Length:	422
Score:	1586.50	Matches:	308
Percent Similarity:	78.91%	Conservative:	25
Best Local Similarity:	72.99%	Mismatches:	34
Query Match:	57.13%	Indels:	56
DB:	11	Gaps:	3

US-09-980-881A-1 (1-1573) x Q9QZF0 (1-422)

```
Qy        18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
Db         1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19

Qy        78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
Db        20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy       138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
Db        40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy       198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
Db        60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79

Qy       258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
Db        80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
Db       100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
```

```
Db        120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValLeuThrGluGlnHis 139

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy        608 ------------------------------------------------------------ 608

Db        200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy        609 -----------------------------------------AATCGAATGTGGAGAAAG 626
                                                      ||||||||||||||||||
Db        220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 279

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              :::   ||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        300 AspHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||         :::||||||||||||||||||||||||   |||||||||
Db        340 ValArgAlaIleGlySerPheAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              :::||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy       1047 TCGTT-------------------------------------------------------TAC 1054
              |||                                                         |||
Db        380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              |||||||||||||||   ||||||   ||||||:::|||||||||   ||||||||||||
Db        400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLysIleValTrpHisValIle 419

Qy       1115 AGGAAT 1120
              ||||||
Db        420 ArgAsn 421

RESULT 6
Q9EQV9
ID   Q9EQV9      PRELIMINARY;      PRT;   422 AA.
AC   Q9EQV9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Pre-procarboxypeptidase R.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20471387; PubMed=11021404;
RA   Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA   Okada N., Okada H.;
RT   "Molecular cloning and partial characterization of rat
RT   procarboxypeptidase R and carboxypeptidase N.";
RL   Microbiol. Immunol. 44:719-728(2000).
DR   EMBL; AB042598; BAB18617.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   422 AA;  48826 MW;  FFFD32A51A9366C8 CRC64;

Alignment Scores:
Pred. No.:              4.29e-149      Length:         422
Score:                  1568.50        Matches:        305
Percent Similarity:     78.44%         Conservative:   26
Best Local Similarity:  72.27%         Mismatches:     36
Query Match:            56.48%         Indels:         56
DB:                     11             Gaps:           3

US-09-980-881A-1 (1-1573) x Q9EQV9 (1-422)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              |||||||||      |||   |||||||||   |||:::|||:::   |||:::|||
Db          1 MetLysLeuTyrGlyLeuGlyValLeuValAlaIleIleLeuTyr---GluLysHisGly 19

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||
Db         20 LeuAlaPheGlnSerGlyHisValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              :::|||||||||||||||||||||||||||:::|||||||||||||||||||||||:::
Db         40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              |||   |||||||||:::|||||||||||||||||||||||||||||:::::|||
Db         60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAsnSerVal 79

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||:::||||||   |||   ||||||   :::|||||:::   :::|||||||||
Db         80 LysAlaTyrLeuAsnAlaSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||   |||||||||||||||||||||||||||:::|||||||||
Db        100 LeuIleGlnGlnGlnThrSerAsnAspThrValSerProArgAlaSerSerSerTyrTyr 119

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              |||||||||||||||||||||||||||||||||||||||||||||   |||||||:::|||
Db        120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              |||||||||||   ||||||:::||||||||||||:::|||||||||||||||||||||
Db        140 ProAspMetLeuGlnLysIleTyrIleGlySerSerTyrGluLysTyrProLeuTyrVal 159

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||          ||||||||||||||||||||||||||||||
Db        160 LeuLysValSerGlyLysGluHisArgValLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy        608 ------------------------------------------------------------ 608

Db        200 PheHisGlyLysGluAsnThrTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219
```

```
Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                      ||||||||||||||||||
Db        220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        240 AsnArgSerValHisMetAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||:::||||||||||||   ||||||||||||||||||||||||
Db        260 SerLysHisTrpCysGluLysGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 279

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              |||   ||||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        300 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              ||||||||||||:::      :::||||||||||||||||||||||||   |||||||||
Db        340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              :::||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy       1047 TCGTTTACATC---------------------------------------------------- 1057
              |||||||||
Db        380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgPhe 399

Qy       1058 ---AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
                 ||||||||||||   ||||||   ||||||||||||||||||||||||||||||||||
Db        400 IleLysProThrCysAlaGluAlaLeuAlaAlaValSerLysIleAlaTrpHisValIle 419

Qy       1115 AGGAAT 1120
              ||||||
Db        420 ArgAsn 421
```

RESULT 7
Q9NTI8
ID Q9NTI8 PRELIMINARY; PRT; 198 AA.
AC Q9NTI8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN CPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tromans A.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL137141; CAB92622.1; -.
DR HSSP; P00730; 1CPX.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
KW Carboxypeptidase.
FT NON_TER 198 198
SQ SEQUENCE 198 AA; 22460 MW; 32F005305621C2A5 CRC64;

```
Alignment Scores:
Pred. No.:              9.09e-95      Length:          198
Score:                  1029.00       Matches:         196
Percent Similarity:     99.49%        Conservative:    0
Best Local Similarity:  99.49%        Mismatches:      1
Query Match:            37.05%        Indels:          0
DB:                     4             Gaps:            0

US-09-980-881A-1 (1-1573) x Q9NTI8 (1-198)

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
```

RESULT 8
Q9PUF2
ID Q9PUF2 PRELIMINARY; PRT; 416 AA.
AC Q9PUF2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase homolog.
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE FROM N.A.
RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT "Screening of Bothrops jararaca pancreas cDNA library.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF190274; AAF01344.1; -.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.

```
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ   SEQUENCE   416 AA;  47723 MW;  4F999854DD72B7A7 CRC64;

Alignment Scores:
Pred. No.:              6.3e-63       Length:         416
Score:                  714.00        Matches:        151
Percent Similarity:     53.94%        Conservative:   68
Best Local Similarity:  37.19%        Mismatches:     127
Query Match:            25.71%        Indels:         61
DB:                     13            Gaps:           6

US-09-980-881A-1 (1-1573) x Q9PUF2 (1-416)

Qy         60 TTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGA 119
Db         14 PheAlaGluThrThrValHisArgPheAspGlyGluLysValTyrArgValThrProArg 33

Qy        120 ACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGG 179
Db         34 AsnGluAspGluValTyrPheLeuAsnTyrLeuAlaAsnIleValGlnValAspPheTrp 53

Qy        180 CAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCA 239
Db         54 ArgProAspSerValGluLeuValLysAlaGluMetThrValAspPheArgIleGluAla 73

Qy        240 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 299
Db         74 AspArgCysSerGluValGluSerIleLeuGlnGlnSerGlyLeuAsnTyrGluIleLeu 93

Qy        300 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 359
Db         94 IleAspAsnLeuGlnAlaValLeuAspArgGlnLeuAspAsnHisAla---------Arg 110

Qy        360 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 419
Db        111 ThrAlaGlyTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAlaTrpThrAla 130

Qy        420 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 479
Db        131 AspIleAlaAsnGluAsnProSerLeuValSerArgLeuGlnIleGlyThrThrPheGlu 150

Qy        480 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 539
Db        151 GlyArgProMetProLeuLeuLysVal---GlyLysProGlyValAsnLysLysAlaIle 169

Qy        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 599
Db        170 PheIleAspCysGlyPheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPhe 189

Qy        600 ATA---------------------------------------------------------- 602
Db        190 ValArgGluAlaValArgThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLys 209

Qy        603 ----------------------------------------------------GGCCAT 608
Db        210 LeuAspPheTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSerTrpLysGln 229

Qy        609 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 668
Db        230 SerArgMetTrpArgLysThrArgSerValAsnAlaGlySerThrCysIleGlyThrAsp 249

Qy        669 CTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 728
Db        250 ProAsnArgAsnPhe---AspAlaAlaTrpCysSerValGlyAlaSerArgAsnProCys 268

Qy        729 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 788
Db        269 SerGluThrTyrCysGlySerLysProGluSerGluLysGluThrLysAlaLeuAlaAsp 288

Qy        789 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 848
Db        289 PheIleArgArgAsnArgSerIleIleGlnAlaTyrLeuThrIleHisSerTyrSerGln 308

Qy        849 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 908
Db        309 MetLeuLeuTyrProTyrSerTyrThrTyrAspLeuThrSerAsnAsnLysLysLeuAsn 328

Qy        909 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACA 968
Db        329 SerIleAlaLysGluAlaIleArgGluLeu---LysValLeuPheGlyThrGluTyrThr 347

Qy        969 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1028
Db        348 TyrGlyProGlyAlaAlaThrIleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyr 367

Qy       1029 GATTTGGGCATCAAATATTCGTTTACATC---------------------------------- 1057
Db        368 AspGlnGlyIleLysTyrAlaPheThrPheGluLeuArgAspLysGlyArgTyrGlyPhe 387

Qy       1058 ----------------------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1096
Db        388 AlaLeuProGluSerGlnIleLysProThrCysGluGluThrMetIleAlaValLysTyr 407

Qy       1097 ATAGCTTGGCATGTCATT 1114
Db        408 IleAlaGluTyrMetLeu 413

RESULT 9
Q8AXN3
ID   Q8AXN3      PRELIMINARY;      PRT;   408 AA.
AC   Q8AXN3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B (Fragment).
GN   CPB.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Srivastava A.S., Kurokawa T., Suzuki T.;
RT   "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT   A2 and B from the Japanese flounder, Paralichthys olivaceus.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB099302; BAC53789.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   408 AA;  45904 MW;  809CAACD9D2DA34B CRC64;

Alignment Scores:
Pred. No.:              7.18e-62      Length:         408
Score:                  703.50        Matches:        153
Percent Similarity:     54.11%        Conservative:   64
Best Local Similarity:  38.15%        Mismatches:     125
Query Match:            25.33%        Indels:         60
DB:                     13            Gaps:           6

US-09-980-881A-1 (1-1573) x Q8AXN3 (1-408)

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
Db         12 PheGluGlyGluLysValPheArgLeuLysProValPheAspGluHisValThrLeuIle 31

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
Db         32 ArgAspLeuAlaAsnSerIleGluValAspPheTrpArgProGluSerSerGlnLeuVal 51
```

```
Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
Db         52 ThrIleAspIleAspValAspIleArgValProAlaIleTyrLeuAspMetValTyrThr 71

Qy        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
Db         72 LeuLeuGlnGlnSerAspMetGluHisGluValLeuIleGluAspLeuGlnSerAlaVal 91

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
Db         92 AspAlaGluAlaAsp---------LeuLysProSerProArgAlaHisSerTyrThrLys 108

Qy        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
Db        109 TyrAsnAsnLeuAspLysValGlnSerTrpIleAlaSerIleSerSerSerAsnProAsp 128

Qy        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
Db        129 LeuIleSerLysGlnValIleGlyAsnThrTyrGluGlyArgProMetThrValLeuLys 148

Qy        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
Db        149 LeuGlyLysLysSerSerSerThrLysProAlaIlePheMetAspCysGlyIleHisAla 168

Qy        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------- 602
Db        169 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValLysGluAlaLeuSerThrTyr 188

Qy        602 ---------- 602

Db        189 GlySerAspSerGlnMetThrSerLeuLeuAsnGlnMetAspValPheValLeuProVal 208

Qy        603 ----------GGCCAT----------AATCGAATGTGGAGAAAGAACCGT 632
Db        209 PheAsnIleAspGlyTyrAspPheThrHisLysSerAsnArgMetTrpArgLysThrArg 228

Qy        633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
Db        229 SerArgLysSerGlySerSerCysIleGlyAlaAspProAsnArgAsnPhe---AspAla 247

Qy        693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
Db        248 GlyTrpCysSerMetGlyAlaSerAspAsnProCysSerAspThrPheCysGlyTyrThr 267

Qy        753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
Db        268 ProGluSerGluIleGluValLysAsnValAlaAspPheIleArgArgAsnArgSerIle 287

Qy        813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
Db        288 IleLysAlaTyrLeuThrValHisSerTyrSerGlnLeuLeuLeuProTyrSerTyr 307

Qy        873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
Db        308 ThrTyrGlnLeuAlaAlaAspHisSerGluLeuMetLysValAlaGluGlyAlaSerAla 327

Qy        933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
Db        328 AlaLeu---ArgSerLeuTyrGlyThrThrTyrThrSerGlyProGlyAlaThrThrIle 346

Qy        993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
Db        347 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspLeuGlyValLysTyrSerTyr 366

Qy       1053 ACATC----------AAA 1060
Db        367 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleLys 386

Qy       1061 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
Db        387 ProThrCysGluGluThrMetLeuAlaValLysTyrIleAlaAlaTyrValGlnLysAsn 406

Qy       1121 GTT 1123
```

```
Db        407 Leu 407
```

RESULT 10
Q9XSP3
ID Q9XSP3 PRELIMINARY; PRT; 416 AA.
AC Q9XSP3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Procarboxypeptidase B precursor (EC 3.4.17.2).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=99321926; PubMed=10391940;
RA Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA Hershberger C., Aviles F.;
RT "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL J. Biol. Chem. 274:19925-19933(1999).
DR EMBL; AJ133775; CAB46991.1; -.
DR HSSP; P09955; 1PBA.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Hydrolase; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 416 PROCARBOXYPEPTIDASE B.
SQ SEQUENCE 416 AA; 47381 MW; 5FA7FCED7B45AE6A CRC64;

Alignment Scores:

Pred. No.:	4.53e-56	Length:	416
Score:	646.00	Matches:	140
Percent Similarity:	53.40%	Conservative:	64
Best Local Similarity:	36.65%	Mismatches:	134
Query Match:	23.26%	Indels:	44
DB:	6	Gaps:	7

US-09-980-881A-1 (1-1573) x Q9XSP3 (1-416)

```
Qy         24 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 83
Db          1 MetLeuAlaPheLeuIleLeuValThrValThrLeuAlaSerAlaHisHisSerGlyGlu 20

Qy         84 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 140
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerLeu 40

Qy        141 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 200
Db         41 LeuHisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        201 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 260
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGlu 80

Qy        261 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 320
Db         81 AspPheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerVal 100

Qy        321 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 380
Db        101 LeuGluAlaGlnPheAspSerArgVal---------ArgThrThrGlyHisSerTyrGlu 117
```

```
Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440
Db        118 LysTyrAsnAsnTrpGluThrIleGluAlaTrpThrLysGlnValThrSerGluAsnPro 137

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500
Db        138 AspLeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeu 157

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560
Db        158 LysVal---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
Db        177 AlaArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThr 196

Qy        602 ------------------------------------------------------------ 602

Db        197 TyrGlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuPro 216

Qy        603 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThr 236

Qy        630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689
Db        237 ArgSerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749
Db        256 AlaGlyTrpCysThrThrGlyAlaSerThrAspProCysAspGluThrTyrCysGlySer 275

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSer 295

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSer 315

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929
Db        316 TyrAspTyrLysLeuProGluAsnAsnAlaGluLeuAsnAsnLeuAlaLysAlaAlaVal 335

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989
Db        336 LysGluLeu---AlaThrLeuTyrGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049
Db        355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrSer 374

Qy       1050 TTTACA 1055
Db        375 PheThr 376
```

RESULT 11
P97597
ID P97597 PRELIMINARY; PRT; 412 AA.
AC P97597;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mast cell carboxypeptidase A precursor (Fragment).
GN R-CPA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague Dawley;
RX MEDLINE=97149430; PubMed=8996238;
RA Lutzelschwab C., Pejler G., Aveskogh M., Hellman L.;
RT "Secretory granule proteases in rat mast cells. Cloning of 10
RT different serine proteases and a carboxypeptidase A from various rat
RT mast cell populations.";
RL J. Exp. Med. 185:13-29(1997).
DR EMBL; U67914; AAB48267.1; -.
DR HSSP; P09955; 1NSA.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Carboxypeptidase; Signal.
FT NON_TER 1 1
FT SIGNAL <1 10 POTENTIAL.
FT CHAIN 11 412 MAST CELL CARBOXYPEPTIDASE A.
SQ SEQUENCE 412 AA; 47944 MW; DAB59555FC49137D CRC64;

Alignment Scores:

Pred. No.:	1.02e-55	Length:	412
Score:	642.50	Matches:	143
Percent Similarity:	51.25%	Conservative:	62
Best Local Similarity:	35.75%	Mismatches:	136
Query Match:	23.14%	Indels:	60
DB:	11	Gaps:	7

US-09-980-881A-1 (1-1573) x P97597 (1-412)

```
Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
Db         16 PheAspArgGluLysValPheArgValLysLeuGlnAspGluLysGlnAlaSerIleLeu 35

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
Db         36 LysAsnLeuThrGlnThrIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIle 55

Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263
Db         56 AlaValAsnMetThrValAspPheArgValThrGluLysGluSerGlnThrIleGlnSer 75

Qy        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323
Db         76 ThrLeuGluGlnHisLysMetAspTyrGluIleLeuIleAsnAspLeuGlnGluGluIle 95

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383
Db         96 AspLysGlnPhe------AspValLysGluGluIleAlaGlyArgHisSerTyrAlaLys 113

Qy        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443
Db        114 TyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLysHisProGlu 133

Qy        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503
Db        134 MetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLys 153

Qy        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563
Db        154 Ile---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAla 172

Qy        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602
Db        173 ArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAlaLysSerTyr 192

Qy        603 GGCCATAAT---------------------------------------------------- 611
Db        193 GlyLysAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProVal 212

Qy        612 -----------------------------------CGAATGTGGAGAAAGAACCGT 632
```

```
Db        213 PheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArgLysAsnArg 232

Qy        633 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 692
              |||          |||:::  ||||||||||||||||||||   ||||||
Db        233 SerLysAsnProAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspVal 251

Qy        693 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 752
                 |||                  :::   :::   |||          |||   |||
Db        252 SerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArgGlyProAla 271

Qy        753 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 812
              ||||||||||||   |||   |||||||||   :::|||:::|||   ::::::|||
Db        272 ProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSer 291

Qy        813 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 872
              ||||||||||||||:::   ||||||||||||||   ::::::|||||||||   |||
Db        292 IleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuProProTyrGlyTyr 311

Qy        873 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 932
              |||                 :::|||::::::|||      ||||||      |||
Db        312 ThrIleLysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAsp 331

Qy        933 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 992
                 :::      :::        |||||||||   :::|||      :::   |||:::
Db        332 ValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIle 350

Qy        993 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1052
              |||          |||          ||||||   |||||||||||||||||||:::|||
Db        351 TyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThrPhe 370

Qy       1053 AC-------------------------------------------------------ATCAAA 1060
                                                                       ||||||
Db        371 AlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLys 390

Qy       1061 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAAT 1120
              |||||||||:::|||         :::|||      ||||||   ::::::::::::::
Db        391 ProThrCysLysGluThrMetLeuSerValLysPheIleAlaLysTyrIleLeuLysHis 410

RESULT 12
Q8N4T0
ID   Q8N4T0      PRELIMINARY;      PRT;   437 AA.
AC   Q8N4T0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033684; AAH33684.1; -.
DR   Genew; HGNC:17245; CPA6.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   437 AA;  51024 MW;  9970680D429A0A98 CRC64;

Alignment Scores:
Pred. No.:              8.36e-55        Length:         437
Score:                  633.50          Matches:        147
Percent Similarity:     49.66%          Conservative:   70
Best Local Similarity:  33.64%          Mismatches:     147
Query Match:            22.81%          Indels:         74
DB:                     4               Gaps:           12

US-09-980-881A-1 (1-1573) x Q8N4T0 (1-437)

Qy          9 GCTGTTGGGATGAAGCTTTGC------AGCCTTGCAGTCCTTGTACCCATTGTTCTCTTC 62
              |||      :::   ||||||         |||   :::|||   |||
Db         12 AlaAlaPheLeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro------------ 27

Qy         63 TGTGAGCAGCATGTCTTC------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCT 116
                    |||::::::         :::         :::|||:::      :::|||
Db         28 GlyHisSerHisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIlePro 47

Qy        117 AGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC 176
              :::|||      :::         |||:::   ::::::         ::::::   |||
Db         48 LysThrGluGluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeu 67

Qy        177 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAG------AAAAAACAAGTCCATTTTTTT 230
              |||||||||   :::         :::   :::                |||||
Db         68 TrpGlnProSerSerIleSerTyrValSerGluGlyThrValThrAspValHisIlePro 87

Qy        231 GTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 290
              |||   |||            :::   |||   |||      :::   |||
Db         88 GlnAsnGlySerArgAla------LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyr 105

Qy        291 AGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 350
              ||||||:::   |||::::::      :::::::::      |||      |||
Db        106 LysValLeuIleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGln 125

Qy        351 AGCCCCCGAGCCTCCGCATCG------TACTATGAACAGTATCACTCACTAAATGAAATC 404
                  |||   |||   |||          ||||||   ||||||||||   ||||||
Db        126 ArgAsnArgArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIle 145

Qy        405 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 464
                 :::|||:::      :::   :::   |||      ::::::          |||
Db        146 GlnAsnTrpMetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIle 165

Qy        465 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA 524
              |||   |||:::|||         |||::::::||||||:::   |||:::
Db        166 GlyArgSerTyrGluGlyArgCysLeuPheIleLeuLysLeu---GlyArgArgSerArg 184

Qy        525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
              |||   |||:::||||||||||||||||||||||||||||||||||||   ||||||
Db        185 LeuLysArgAlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAla 204

Qy        585 TTCTGCTTGTGGTTCATA---------------------------------------- 602
              ||||||   ||||||:::
Db        205 PheCysGlnTrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArg 224

Qy        602 ---------------------------------------------------------- 602

Db        225 LysMetLeuAsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHis 244

Qy        603 ---------GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 653
                      ::::::|||   |||||||||   ||||||      :::
Db        245 PheSerTrpThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArg 264

Qy        654 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 713
              |||   |||   |||   |||   |||:::      |||   ||||||:::|||||||||
Db        265 CysArgGlyValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAla 283

Qy        714 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 773
              |||         |||   :::|||||||||||   :::|||||||||||||||||||
Db        284 SerMetHisProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluVal 303

Qy        774 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 833
              ||||||||||||:::|||||||||:::::          |||:::||||||:::|||
Db        304 LysAlaValAlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPhe 323

Qy        834 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 893
              |||:::|||:::|||   :::::::::|||||||||||       :::         :::
Db        324 HisAlaTyrAlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsn 343

Qy        894 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 953
                    :::         |||   :::||||||   |||:::   ::::::
```

```
Db        344 PheArgCysValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyr 362

Qy        954 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 1013
                  ||||||   :::|||      |||   |||||||||::::::   |||
Db        363 GlyValArgTyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSer 382

Qy       1014 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC------------------- 1054
                  ||||||   |||      ||||||   |||:::|||
Db        383 MetAspTrpAlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThr 402

Qy       1055 --------------------------------ATCAAACCCACCTGTAGAGAAGCTTTT 1081
                                               ||||||||||||||   |||
Db        403 GlyTyrPheGlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMet 422

Qy       1082 GCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCC 1132
                  ||||||      |||      |||:::::::::         ||||||
Db        423 LeuAlaValLysAsnIleThrMetHisLeuLeuLysLys------CysPro 437

RESULT 13
Q8AVS2
ID   Q8AVS2      PRELIMINARY;      PRT;   434 AA.
AC   Q8AVS2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041315; AAH41315.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   434 AA;  50581 MW;  B42BD77843910591 CRC64;

Alignment Scores:
Pred. No.:              6.01e-54        Length:         434
Score:                  625.00          Matches:        147
Percent Similarity:     49.30%          Conservative:   65
Best Local Similarity:  34.19%          Mismatches:     139
Query Match:            22.51%          Indels:         80
DB:                     13              Gaps:           11

US-09-980-881A-1 (1-1573) x Q8AVS2 (1-434)

Qy         39 GTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAA 98
              ||||||:::   :::::::|||   |||               |||      :::||||||
Db         11 ValLeuLeuPheLeuLeuLeuSerCys--------------AlaLeuSerAlaGlyGln 25

Qy         99 ----------------------------------GTTCTAGCTGCTCTTCCTAGAACC 122
                                                   |||:::         |||:::
Db         26 CysHisLeuTyrAsnAsnArgTyrSerTrpAspLysValIleArgValPheProLysAsn 45

Qy        123 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAG 182
                  |||:::      |||:::::::::          ::::::   |||||||||
Db         46 GluArgGluAlaAsnGluLeuLysAspMetTyrGlnGlnLeuGlnValAspLeuTrpGln 65

Qy        183 CCGGTAACAGCTGACCTTATTGTGAAGAAA------AAACAAGTCCATTTTTTTGTAAAT 236
                  |||   :::         |||   |||               ||||||         |||
Db         66 ProSerSerIleSerHisIleGlyLysAspThrValThrAspValHisThrSerGlyAsn 85

Qy        237 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 296
                  :::|||:::|||      :::      :::|||      :::   |||         :::
Db         86 SerSerGluVal------LeuLeuThrTyrLeuThrLysAlaLysIleGlnHisLysIle 103

Qy        297 TTGCTGGCAGACGTGGAAGATCTTATTCAA---CAGCAGATTTCCAACGACACAGTCAGC 353
                  |||:::   :::|||:::   :::::::::   ||||||
Db        104 LeuValAsnAsnValGlnSerMetLeuGluAlaGlnGlnAlaPheArgProArgArgLys 123

Qy        354 CCCCGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCT 410
                  |||:::   :::   |||   |||   :::||||||   |||:::||||||   |||
Db        124 ArgArgSerLeuSerArgTyrAsnTyrAsnGluTyrHisProLeuHisGluIleGluSer 143

Qy        411 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 470
                  |||:::   ::::::   :::   |||   |||:::::::::         ||||||
Db        144 TrpMetPheTyrMetAsnLysThrHisHisAspLeuValSerLeuPheThrIleGlyLys 163

Qy        471 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAA 530
                  |||:::|||          |||||||||||||||:::             ::::::   |||
Db        164 SerTyrGluGlyArgSerLeuTyrValLeuLysLeuGlyLysAspThrLysSerTyrLys 183

Qy        531 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGC 590
                  |||||||||||||||||||:::|||||||||||||||||   ||||||||||||
Db        184 LysAlaIleTrpIleAspCysGlyMetHisAlaArgGluTrpIleGlyProAlaPheCys 203

Qy        591 TTGTGGTTCATA---------------------------------------------- 602
                  ||||||:::
Db        204 GlnTrpPheValLysGluAlaIleAsnSerTyrAsnThrAspProAlaMetLysLysIle 223

Qy        603 ---------------------------------------------GGCCAT--------- 608
                                                             |||:::
Db        224 LeuAsnLeuLeuTyrIleTyrValMetProValPheAsnValAspGlyTyrHisTyrSer 243

Qy        609 ---------AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 659
                         :::|||   |||||||||   ||||||                  |||
Db        244 TrpHisSerAspArgPheTrpArgLysThrArgSerLysAsnThrArgTyrGlnCysTyr 263

Qy        660 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 719
                  |||   |||   |||   |||:::         ||||||   :::|||||||||||
Db        264 GlyValAspAlaAsnArgAsnTrp---LysValHisTrpSerAspGluGlyAlaSerLeu 282

Qy        720 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 779
                  :::   |||      |||||||||||   |||   |||||||||||||||||||||
Db        283 AsnProCysAspAsnThrTyrCysGlyProTyrAlaGluSerGluProGluValLysAla 302

Qy        780 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 839
                  ||||||   ||||||   :::            ::::::||||||:::|||   |||:::
Db        303 ValAlaGlnPheLeuTyrLysGlnArgLysHisValArgAlaTyrMetSerPheHisAla 322

Qy        840 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 899
                  |||:::|||   :::::::::||||||||||||                  :::
Db        323 TyrAlaGlnMetLeuLeuTyrProTyrSerTyrGlnTyrGlyAlaIleProAsnPheGly 342

Qy        900 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959
                     :::         |||      ||||||   ||||||   :::::::::
Db        343 CysValGluSerAlaAlaHisAsnAlaValLeuAlaIle---ArgSerAlaTyrGlyIle 361

Qy        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019
                  ||||||   ||||||      |||   ||||||||||||      |||         |||
Db        362 ArgTyrArgHisGlyProAlaSerSerThrLeuTyrLeuThrSerGlySerSerMetAsp 381

Qy       1020 TGGATCTATGATTTGGGCATCAAATATTCGTTTAC------------------------- 1054
                  |||   |||:::   ||||||   ||||||:::
Db        382 TrpAlaTyrAsnAsnGlyIleProTyrSerTyrAlaPheGluLeuArgAspThrGlyTyr 401

Qy       1055 --------------------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 1087
                                          ||||||||||||||   |||         |||
Db        402 TyrGlyPheLeuLeuProGluGlyLeuIleLysProThrCysValGluThrMetLeuAla 421

Qy       1088 GTCTCTAAAATAGCTTGGCATGTCATTAGG 1117
                  |||      |||      |||   ::::::
Db        422 ValLysAsnIleThrMetHisAlaLeuLys 431

RESULT 14
Q8UUK1
ID   Q8UUK1      PRELIMINARY;      PRT;   419 AA.
AC   Q8UUK1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Preprocarboxypeptidase A precursor.
```

```
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=domesticus; TISSUE=Pancreas;
RA   Hasegawa S., Honda K., Hikami Y.;
RT   "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
RT   precursor.";
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X64539; CAA45837.1; -.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Signal; Carboxypeptidase.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN       113    419       CARBOXYPEPTIDASE A.
SQ   SEQUENCE   419 AA;  47006 MW;  299C334274BCBC9C CRC64;

Alignment Scores:
Pred. No.:              1.63e-49      Length:        419
Score:                  581.00        Matches:       133
Percent Similarity:     52.52%        Conservative:  65
Best Local Similarity:  35.28%        Mismatches:    107
Query Match:            20.92%        Indels:        72
DB:                     13            Gaps:          9

US-09-980-881A-1 (1-1573) x Q8UUK1 (1-419)

Qy         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143
              |||         |||||||||      :::|||   :::      :::::::|||   :::
Db         21 PheValGlyHisGlnValLeuArgIleValProSerSerAspAlaGluLeuGlnLysVal 40

Qy        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203
              |||   |||
Db         41 GlnGluLeuGlnGlu---------------------------------------------- 45

Qy        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT--------- 254
              ::::::      |||:::   |||:::::::::        :::   |||
Db         46 LeuGluHisLeuGlnLeuAspPheTrpLeuSerProArgGlyLeuGlyAsnProValAsp 65

Qy        255 ----------------------------GTGAAAGCCCATTTAAATGTGAGCGGAATT 284
                                          ||||||||||||||||      :::|||:::
Db         66 IleArgValProPheProSerLeuGlnProValLysAlaHisLeuGluAlaAsnGlyVal 85

Qy        285 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT------------CAACAGCAG 332
              |||   |||:::::::::   ||||||:::   |||:::               |||
Db         86 ProTyrSerIleMetIleGluAspValGlnAlaLeuValAspLeuGluGlnMetGlnMet 105

Qy        333 ATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAACAGTATCAC 389
              :::                 |||   ::::::::::::::   |||      ||||||
Db        106 LeuArgArgArgArgPheValProLeuSerThrSerThrPheAspTyrThrSerTyrHis 125

Qy        390 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 449
              :::|||   |||||||||:::::::::::   :::         :::|||::::::::
Db        126 ThrLeuGluGluIleTyrAlaPheMetAspLeuLeuValAlaGluAsnProAsnLeuVal 145

Qy        450 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 509
              :::|||:::   ||||||   :::   |||      |||:::|||||||||||   |||
Db        146 SerLysLeuGluIleGlyArgThrThrGluAsnArgProIleTyrValLeuLysPheSer 165

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
                 |||      |||   :::   |||||||||||||   |||||||||:::||||||
Db        166 ---LysGlyGlyThrAsnArgProAlaIleTrpIleAspThrGlyIleHisSerArgGlu 184

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 602
              |||::::::   |||      :::||||||
Db        185 TrpValThrGlnAlaSerGlyIleTrpPheAlaLysLysIleValGlnGluGlnAspGlu 204

Qy        602 ---------------------------------------------------------- 602

Db        205 GlyLeuAlaAsnIleLeuAspGlnMetAspIlePheLeuGluIleValThrAsnProAsp 224

Qy        603 ------------------GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCG 644
                                |||||||||||||||||   ||||||      :::
Db        225 GlyPheAlaPheThrHisThrSerAsnArgMetTrpArgLysThrArgSerLysArgSer 244

Qy        645 AACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAG 704
                 :::   |||:::|||   |||   |||   |||:::           :::
Db        245 GlySerLeuCysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGly 263

Qy        705 GAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAA 764
                |||||||||||:::   |||||||||||||||   |||   |||      ||||||
Db        264 SerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyProTyrAlaAsnSerGlu 283

Qy        765 CCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATAC 824
              |||||||||||||:::      |||::::::   :::   |||   |||||||||:::
Db        284 ProGluValLysAlaIleValAspPheValLysSerHisLysAsn---IleLysAlaPhe 302

Qy        825 ATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAA 884
              ||||||:::|||||||||||||||   :::::::::||||||   ||||||   :::
Db        303 IleSerIleHisSerTyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrThrAla 322

Qy        885 AGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAA 944
                 |||   |||||||||      ||||||   |||||||||   |||:::
Db        323 ValProAspLysGluGluLeuHisGlnValAlaLysGluAlaValAlaAlaLeu---Ser 341

Qy        945 ACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCT 1004
              :::          |||   |||   :::|||            |||:::|||   |||
Db        342 SerLeuTyrGlyThrAsnTyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSer 361

Qy       1005 GGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1055
              ||||||      ||||||   |||:::   ||||||||||||||||||||
Db        362 GlyGlyThrIleAspTrpThrTyrAsnGlnGlyIleLysTyrSerPheThr 378

RESULT 15
Q8IVL8
ID   Q8IVL8      PRELIMINARY;      PRT;   374 AA.
AC   Q8IVL8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zn-carboxypeptidase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Obaya A.J., Lopez-Otin C.;
RT   "A new Zn-carboxypeptidase highly expressed in ovary.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ422118; CAD19478.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   374 AA;  42529 MW;  404C373BB841AAD2 CRC64;

Alignment Scores:
Pred. No.:              2.04e-48      Length:        374
Score:                  570.00        Matches:       117
Percent Similarity:     52.47%        Conservative:  53
Best Local Similarity:  36.11%        Mismatches:    90
Query Match:            20.53%        Indels:        65
DB:                     4             Gaps:          7

US-09-980-881A-1 (1-1573) x Q8IVL8 (1-374)

Qy        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAA 380
```

```
Db         29 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 48

Qy        381 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 440

Db         49 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 68

Qy        441 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 500

Db         69 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu 88

Qy        501 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 560

Db         89 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 108

Qy        561 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 602

Db        109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128

Qy        602 ------------------------------------------------------------ 602

Db        129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 148

Qy        603 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 629

Db        149 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 168

Qy        630 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 689

Db        169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187

Qy        690 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 749

Db        188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207

Qy        750 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 809

Db        208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227

Qy        810 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 869

Db        228 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247

Qy        870 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 929

Db        248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267

Qy        930 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 989

Db        268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 286

Qy        990 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1049

Db        287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 306

Qy       1050 TTTACATC-------------------------------------------------- 1057

Db        307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326

Qy       1058 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCT--------------- 1102

Db        327 GlnProThrCysGluGluThrMetGluAlaValLeuSerValLeuAspAspValTyrAla 346

Qy       1103 ------TGGCAT 1108

Db        347 LysHisTrpHis 350


Search completed: January  1, 2004, 18:58:33
Job time : 106.358 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:58:24 ; Search time 5920.84 Seconds
(without alignments)
10868.540 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1528	97.1	1546	9	AB011969	AB011969 Homo sapi
2		1469.8	93.4	1625	6	AR232183	AR232183 Sequence
3		1372	87.2	1728	6	AX409750	AX409750 Sequence
4		1372	87.2	1728	6	AX706330	AX706330 Sequence
5		1372	87.2	1728	9	HUMPCPBX	M75106 Human prepr
6		1372	87.2	1749	6	I16100	I16100 Sequence 2
7		1372	87.2	1749	6	I33526	I33526 Sequence 2
8		1309	83.2	1715	9	BC007057	BC007057 Homo sapi
9		918	58.4	1272	6	AR086324	AR086324 Sequence
10		918	58.4	1272	6	BD106653	BD106653 Method of
11		915.4	58.2	1272	9	BT006936	BT006936 Homo sapi
12		914.4	58.1	1272	12	BT007990	BT007990 Synthetic
13		656	41.7	1421	10	AB021968	AB021968 Mus muscu
14		656	41.7	1490	10	AF164524	AF164524 Mus muscu
15		654	41.6	1430	10	AF186188	AF186188 Mus muscu
16		652.6	41.5	1425	10	AB042598	AB042598 Rattus no
17		478.6	30.4	55827	6	AR232184	AR232184 Sequence
18	c	475.4	30.2	58097	9	AL157758	AL157758 Human DNA
19		409	26.0	437	11	G05971	G05971 human STS W
20	c	173.6	11.0	181	11	G59992	G59992 SHGC-130613
21	c	156	9.9	251806	2	AC096305	AC096305 Rattus no
22		156	9.9	258265	2	AC120472	AC120472 Rattus no
23	c	137.2	8.7	431	6	AX400792	AX400792 Sequence
24		128	8.1	180680	9	AL137141	AL137141 Human DNA
25		124	7.9	927	6	AR072911	AR072911 Sequence
26		123.8	7.9	1327	5	AF190274	AF190274 Bothrops
27		123.4	7.8	921	6	I67699	I67699 Sequence 3
28		123.4	7.8	1215	6	I67698	I67698 Sequence 1
29		123.4	7.8	1251	4	SSC133775	AJ133775 Sus scrof
30	c	123.2	7.8	451	11	G53326	G53326 SHGC-82320
31		120.8	7.7	1251	4	DOGZAP47	D78348 Canis famil
32		109	6.9	1263	6	A87530	A87530 Sequence 11
33		107.6	6.8	1622	9	HUMCARMC	M27717 Human mast
34		107.6	6.8	1622	11	G28614	G28614 human STS S
35		107.6	6.8	1674	9	BC012613	BC012613 Homo sapi
36		107.4	6.8	999	6	A51908	A51908 Sequence 72
37		107.4	6.8	999	6	AR085871	AR085871 Sequence
38		107.4	6.8	1053	6	A51904	A51904 Sequence 68
39		107.4	6.8	1053	6	AR085868	AR085868 Sequence
40		107.4	6.8	1254	12	BT008111	BT008111 Synthetic
41		107.4	6.8	1263	6	A51896	A51896 Sequence 60
42		107.4	6.8	1263	6	AR085861	AR085861 Sequence
43		107.4	6.8	1284	6	A51913	A51913 Sequence 77
44		107.4	6.8	1284	6	AR085875	AR085875 Sequence
45		107.4	6.8	1310	9	HSA224866	AJ224866 Homo sapi

ALIGNMENTS

RESULT 1
AB011969
LOCUS AB011969 1546 bp mRNA linear PRI 02-FEB-2000
DEFINITION Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.
ACCESSION AB011969
VERSION AB011969.1 GI:6855463
KEYWORDS carboxypeptidase B-like protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE Isolation, molecular cloning, and partial characterization of a

```
          novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
REFERENCE   2  (sites)
  AUTHORS   Matsumoto,A.
  TITLE     Isolation, molecular cloning, and partial characterization of a
          novel carboxypeptidase B from human plasma
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1546)
  AUTHORS   Matsumoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
          Medicine, Department of Radiation Biophysics & Genetics;
          Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
          (E-mail:amat@med.kobe-u.ac.jp, Tel:81-78341-7451)
COMMENT     Sequence updated (10-Apr-1998).
FEATURES             Location/Qualifiers
     source          1. .1546
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             20. .1102
                     /note="carboxypeptidase B-like protein in human brain"
                     /codon_start=1
                     /product="carboxypeptidase B-like protein"
                     /protein_id="BAA90475.1"
                     /db_xref="GI:6855464"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQTGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDL
                     NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYS
                     QHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDD
                     WIYDLGIKYSFTSNPPVEKLLPLSLK"
     polyA_site      1546
                     /note="21 A nucleotides"
BASE COUNT      453 a    323 c    297 g    473 t
ORIGIN

  Query Match           97.1%;  Score 1528;  DB 9;  Length 1546;
  Best Local Similarity   99.4%;  Pred. No. 0;
  Matches 1534;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| |||| ||||||||||||||||||||||||||||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGACTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 122

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TGGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        483 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATAT 542

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        543 GGATTGACTGTGGAATTCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 602

Qy        601 TAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCG 662

Qy        661 GAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTT 720
              |||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db        663 GAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTT 722

Qy        721 CCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 CCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAG 782

Qy        781 TGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCAT 842

Qy        841 ACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 ACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGG 902

Qy        901 AACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCA 960
              |||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db        903 AACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCA 962

Qy        961 GGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 GGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATT 1022

Qy       1021 GGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1023 GGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTT 1082

Qy       1081 TGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 TGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTA 1142

Qy       1141 TCATTCTGCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1143 TCATTCTGCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT 1202

Qy       1201 TATTTTTAAGTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 TATTTTTAAGTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAG 1262

Qy       1261 AGAACCTAATAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 AGAACCTAATAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCT 1322

Qy       1321 TTAAAAAAAAATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 TTAAAAAAAAATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACT 1382

Qy       1381 AGCCATCTCAAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 AGCCATCTCAAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGAT 1442

Qy       1441 CATCTCACGCTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 CATCTCACGCTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTA 1502

Qy       1501 AAGATTTCTGCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTT 1544
              |||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1503 AAGATTTCTGCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTT 1546


RESULT 2
AR232183
LOCUS       AR232183               1625 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6455294.
ACCESSION   AR232183
VERSION     AR232183.1  GI:27274083
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1625)
  AUTHORS   Gan,W., Ketchum,K.A., Di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human protease proteins, nucleic acid molecules encoding
            human protease proteins, and uses thereof
  JOURNAL   Patent: US 6455294-A 1 24-SEP-2002;
FEATURES             Location/Qualifiers
     source          1. .1625
                     /organism="unknown"
BASE COUNT      488 a    334 c    315 g    488 t
ORIGIN

  Query Match             93.4%;  Score 1469.8;  DB 6;  Length 1625;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches 1553;  Conservative    0;  Mismatches   17;  Indels   53;  Gaps    2;

Qy          2 GAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTT 60

Qy         62 CTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAAC 121
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         61 CTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAAC 120

Qy        122 CTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCA 180

Qy        182 GCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATC 240

Qy        242 TGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCT 300

Qy        302 GGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGC 361
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGC 360

Qy        362 CTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATT 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATT 420

Qy        422 TATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAA 480

Qy        482 GTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATG 541
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        481 GTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATG 540

Qy        542 GATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCAT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCAT 600

Qy        602 AGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGG 660

Qy        662 AACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTC 721
```

```
              |||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||
Db        661 AACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTC 720

Qy        722 CTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGT 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGT 780

Qy        782 GGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATA 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATA 840

Qy        842 CTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGA 900

Qy        902 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAG 961
              |||||||||||||||||||||||||||||||||||| |||| ||||||||||||||||
Db        901 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAG 960

Qy        962 GTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTG 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTG 1020

Qy       1022 GATCTATGATTTGGGCATCAAATATTCGT------------------------------ 1050
              ||||||||||||||||||||||||||||
Db       1021 GATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATA 1080

Qy       1051 ----------------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGT 1089
                                     ||||||||||||||||||||||||||||||||||||
Db       1081 CGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGT 1140

Qy       1090 CTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGC 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGC 1200

Qy       1150 TTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAA 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAA 1260

Qy       1210 GTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAA 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAA 1320

Qy       1270 TAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAA 1329
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1321 TAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAA 1379

Qy       1330 AATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTC 1389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 AATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTC 1439

Qy       1390 AAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACG 1449
              |||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1440 AAGCAAGTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAATGATCATCTCACG 1499

Qy       1450 CTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCT 1509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 CTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCT 1559

Qy       1510 GCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAAAAAAAA 1569
              ||||||||||||||||||||| ||| | | || |||||||||| ||||||||||||
Db       1560 GCTCCAAATTTTCAATAAATTTCTGCTTGTGCCTTTAAAAAAAAAAATAAAAAAAAAAAA 1619

Qy       1570 AAA 1572
              |||
Db       1620 ATA 1622


RESULT 3
AX409750
LOCUS       AX409750               1728 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2397 from Patent WO0229103.
ACCESSION   AX409750
```

```
VERSION     AX409750.1  GI:21442455
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2397 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M75106"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

  Query Match             87.2%;  Score 1372;  DB 6;  Length 1728;
  Best Local Similarity   90.3%;  Pred. No. 2.6e-306;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 122

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 542

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 602

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        603 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 662

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        663 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 722

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 782

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 842

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 902

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 962

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1022

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1023 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1082

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1142

Qy       1030 ATTTGGGCATCAAATATTCGT------------------------------------- 1050
              |||||||||||||||||||||
Db       1143 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1202

Qy       1051 -------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1322

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1382

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1442

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1502

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1562

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1622

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1682

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAA 1563
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1683 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAA 1728


RESULT 4
AX706330
LOCUS       AX706330               1728 bp    DNA     linear   PAT 04-APR-2003
DEFINITION  Sequence 1 from Patent WO03014391.
ACCESSION   AX706330
```

```
VERSION     AX706330.1  GI:29562754
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Morten,J.E. and Jackson,K.N.
  TITLE     Method
  JOURNAL   Patent: WO 03014391-A 1 20-FEB-2003;
            AstraZeneca AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

  Query Match             87.2%;  Score 1372;  DB 6;  Length 1728;
  Best Local Similarity   90.3%;  Pred. No. 2.6e-306;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 122

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 542

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 602

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        603 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 662

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        663 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 722

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        723 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 782

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 842

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 902

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 962

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1022

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1023 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1082

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1142

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||||
Db       1143 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1202

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1322

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1382

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1442

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1502

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1562

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1622

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1682

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAA 1563
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1683 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAA 1728


RESULT 5
HUMPCPBX
LOCUS       HUMPCPBX               1728 bp    mRNA    linear   PRI 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1  GI:189686
```

```
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1728)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
COMMENT     Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     gene            1. .1728
                     /gene="pCPB"
     CDS             20. .1291
                     /gene="pCPB"
                     /note="prepro-plasma carboxypeptidase B"
                     /codon_start=1
                     /protein_id="AAA60042.1"
                     /db_xref="GI:189687"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
     mat_peptide     86. .1288
                     /gene="pCPB"
                     /product="plasma carboxypeptidase B"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN
```

Query Match 87.2%; Score 1372; DB 9; Length 1728;
Best Local Similarity 90.3%; Pred. No. 2.6e-306;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 122

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 542

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 602

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        603 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 662

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        663 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 722

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 782

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 842

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 902

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 962

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1022

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1023 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1082

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1142

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||||
Db       1143 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1202

Qy       1051 -------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1322

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1382

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1442

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1502

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
```

```
Db      1503 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1562

Qy      1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1563 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1622

Qy      1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1623 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1682

Qy      1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAA 1563
             ||||||||||||||||||||||||||||||||||||||||||||
Db      1683 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAA 1728


RESULT 6
I16100
LOCUS       I16100                  1749 bp    DNA     linear   PAT 03-APR-1996
DEFINITION  Sequence 2 from patent US 5474901.
ACCESSION   I16100
VERSION     I16100.1  GI:1251008
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Antibodies to human carboxypeptidase B and methods of use thereof
  JOURNAL   Patent: US 5474901-A 2 12-DEC-1995;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

  Query Match             87.2%;  Score 1372;  DB 6;  Length 1749;
  Best Local Similarity   90.3%;  Pred. No. 2.6e-306;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT-------------------------------------------------------- 608
              ||||||||
Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ------------------------------------------------------------A 609
                                                                           |
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749


RESULT 7
I33526
LOCUS       I33526                1749 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 2 from patent US 5593674.
ACCESSION   I33526
VERSION     I33526.1  GI:1824317
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Plasma carboxypeptidase
  JOURNAL   Patent: US 5593674-A 2 14-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

  Query Match             87.2%;  Score 1372;  DB 6;  Length 1749;
  Best Local Similarity   90.3%;  Pred. No. 2.6e-306;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy         601 TAGGCCAT---------------------------------------------------- 608
               ||||||||
Db         624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy         609 ----------------------------------------------------------A 609
                                                                           |
Db         684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy         610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy         670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
               ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db         804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy         730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy         790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy         850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy         910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy         970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy        1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
               |||||||||||||||||||
Db        1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy        1051 -------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db        1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy        1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy        1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy        1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy        1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy        1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy        1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy        1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703
```

```
Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749


RESULT 8
BC007057
LOCUS       BC007057               1715 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
            IMAGE:3934520, mRNA, complete cds.
ACCESSION   BC007057
VERSION     BC007057.1  GI:13937896
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1715)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 16 Row: f Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4503004.
FEATURES             Location/Qualifiers
     source          1. .1715
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:1361"
                     /db_xref="taxon:9606"
                     /clone="MGC:12495 IMAGE:3934520"
                     /tissue_type="Skeletal Muscle"
                     /clone_lib="NIH_MGC_81"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             25. .1296
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma)"
                     /protein_id="AAH07057.1"
                     /db_xref="GI:13937897"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      518 a    348 c    341 g    508 t
ORIGIN

  Query Match             83.2%;  Score 1309;  DB 9;  Length 1715;
  Best Local Similarity   89.4%;  Pred. No. 9.4e-292;
  Matches 1512;  Conservative    0;  Mismatches   15;  Indels  164;  Gaps    3;
```

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 67

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 127

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 187

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        188 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 247

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        248 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 307

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        308 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 367

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 427

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 487

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        488 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATAT 547

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        548 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 607

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        608 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 667

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        668 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGA 727

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        728 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 787

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        788 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 847

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 907

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 967

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        968 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1027

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db       1028 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1087
```

```
Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1088 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1147

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||||
Db       1148 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1207

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1208 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1267

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1268 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1327

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1328 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1387

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1388 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1447

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1448 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAAAATTGTAA 1506

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1507 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1566

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              |||||||||||||||||||||||||||||| ||||| ||||||||||||||||||||||
Db       1567 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTGATCAATGATCATCTCACGCTGATCAT 1626

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db       1627 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCATA 1686

Qy       1518 TTTTCAATAAA 1528
                 || |||
Db       1687 AAAAAAAAAAA 1697


RESULT 9
AR086324
LOCUS       AR086324               1272 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 1 from patent US 5985562.
ACCESSION   AR086324
VERSION     AR086324.1  GI:10013090
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.John. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk associated with plasma
            carboxypeptidase B polymorphisms
  JOURNAL   Patent: US 5985562-A 1 16-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="unknown"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN

  Query Match           58.4%;  Score 918;  DB 6;  Length 1272;
  Best Local Similarity   86.8%;  Pred. No. 2e-201;
  Matches 1104;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        609 ------------------------------------------------------------ 608

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        609 ------------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140
```

```
Qy       1047 TCGT----------------------------------------------------TTAC 1054
              ||||                                                    ||||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy       1115 AGGAATGTTTAA 1126
              ||||||||||||
Db       1261 AGGAATGTTTAA 1272
```

RESULT 10
BD106653

LOCUS	BD106653 1272 bp DNA linear PAT 18-SEP-2002
DEFINITION	Method of detecting thrombotic disease risk.
ACCESSION	BD106653
VERSION	BD106653.1 GI:23201471
KEYWORDS	JP 2002502253-A/1.
SOURCE	synthetic construct
ORGANISM	synthetic construct artificial sequences.
REFERENCE	1 (bases 1 to 1272)
AUTHORS	Morser,M.J. and Nagashima,M.
TITLE	Method of detecting thrombotic disease risk
JOURNAL	Patent: JP 2002502253-A 1 22-JAN-2002; SCHERING AG
COMMENT	PN JP 2002502253-A/1 PD 22-JAN-2002 PF 02-JUN-1998 JP 1999501465 PR 03-JUN-1997 US 08/869057 PI MICHAEL JOHN MORSER,MARIKO NAGASHIMA PC C12Q1/68 CC Strandedness: Double; CC Topology: Linear; FH Key Location/Qualifiers.
FEATURES	Location/Qualifiers
source	1. .1272 /organism="synthetic construct" /mol_type="genomic DNA" /db_xref="taxon:32630"
BASE COUNT	375 a 269 c 271 g 357 t
ORIGIN	

Query Match 58.4%; Score 918; DB 6; Length 1272;
Best Local Similarity 86.8%; Pred. No. 2e-201;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

```
Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        609 ------------------------------------------------------------ 608

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1047 TCGT----------------------------------------------------TTAC 1054
              ||||                                                    ||||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy       1115 AGGAATGTTTAA 1126
              ||||||||||||
Db       1261 AGGAATGTTTAA 1272
```

RESULT 11
BT006936

LOCUS	BT006936 1272 bp mRNA linear PRI 13-MAY-2003
DEFINITION	Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA, complete cds.
ACCESSION	BT006936
VERSION	BT006936.1 GI:30582710
KEYWORDS	FLI_CDNA.

```
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00156X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .1272
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma, carboxypeptidase
                     U)"
                     /protein_id="AAP35582.1"
                     /db_xref="GI:30582711"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      373 a    268 c    273 g    358 t
ORIGIN

  Query Match             58.2%;  Score 915.4;  DB 9;  Length 1272;
  Best Local Similarity   86.7%;  Pred. No. 8.1e-201;
  Matches 1102;  Conservative    0;  Mismatches    6;  Indels  163;  Gaps    2;

Qy          18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy          78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
               |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy         318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy         378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy         438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy         498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
               ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         481 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy         558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy         609 ------------------------------------------------------------ 608

Db         601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy         609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db         661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy         627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
               |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db         721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy         687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy         747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy         807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy         867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy         927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
               |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy         987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        1047 TCGT----------------------------------------------------TTAC 1054
               ||||                                                    ||||
Db        1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy        1115 AGGAATGTTTA 1125
               ||||||||||| 
Db        1261 AGGAATGTTTA 1271
```

```
RESULT 12
BT007990
LOCUS       BT007990               1272 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
            carboxypeptidase U) mRNA, partial cds.
ACCESSION   BT007990
VERSION     BT007990.1  GI:30584818
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00156L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .>1272
                     /note="Mutations: 1271:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens carboxypeptidase B2 (plasma,
                     carboxypeptidase U)"
                     /protein_id="AAP36662.1"
                     /db_xref="GI:30584819"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNVL"
BASE COUNT      372 a    268 c    273 g    359 t
ORIGIN

  Query Match           58.1%;  Score 914.4;  DB 12;  Length 1272;
  Best Local Similarity  86.7%;  Pred. No. 1.4e-200;
  Matches 1101;  Conservative    0;  Mismatches    6;  Indels  163;  Gaps    2;

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        609 ------------------------------------------------------------ 608

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        609 ------------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140
```

```
Qy       1047 TCGT-------------------------------------------------------TTAC 1054
                ||||                                                       ||||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy       1115 AGGAATGTTT 1124
                ||||||||||
Db       1261 AGGAATGTTT 1270
```

```
RESULT 13
AB021968
LOCUS       AB021968               1421 bp    mRNA    linear   ROD 27-JUL-2000
DEFINITION  Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION   AB021968
VERSION     AB021968.1  GI:9558447
KEYWORDS    carboxypeptidase R.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Obata,K., Okada,N.,
            Campbell,W. and Okada,H.
  TITLE     Pro-carboxypeptidase R is an acute phase protein in the mouse,
            whereas carboxypeptidase N is not
  JOURNAL   J. Immunol. 165 (2), 1053-1058 (2000)
  MEDLINE   20341711
   PUBMED   10878383
REFERENCE   2  (bases 1 to 1421)
  AUTHORS   Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1998) Tomoo Sato, Fukushimura Hospital, Choju
            Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
            441-8124, Japan (E-mail:tomoo@tcp-ip.or.jp, Tel:81-532-46-7511,
            Fax:81-532-46-4899)
FEATURES             Location/Qualifiers
     source          1. .1421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS             17. .1285
                     /codon_start=1
                     /product="carboxypeptidase R"
                     /protein_id="BAB03402.1"
                     /db_xref="GI:9558448"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
     polyA_site      1421
                     /note="11 a nucleotides"
BASE COUNT      418 a    320 c    305 g    378 t
ORIGIN
```

```
  Query Match             41.7%;  Score 656;  DB 10;  Length 1421;
  Best Local Similarity   72.2%;  Pred. No. 6.7e-141;
  Matches 1009;  Conservative    0;  Mismatches  215;  Indels  174;  Gaps    5;

Qy          4 AAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCT 63
              ||  |||||||||||||||||     |||||| | |||| ||| |||| | ||||
Db          3 AAGTCACTGTTGGGATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTA-- 60

Qy         64 GTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCT 123
              |||||||||| |||||| || |||||||||| ||| || |||||||||| |||||||
Db         61 -TGAGCAGCATGGCTTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCT 119

Qy        124 CTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGC 183
              | |||||||||||| | || |||||||||||||||||| |||||| | ||||||||||||
Db        120 CCAGGCAAGTTCAACTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGC 179

Qy        184 CGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTG 243
              | || ||||||||  | || | |||||||||  ||||||| |||||||| ||||| ||||
Db        180 CAGTGACAGCTGAATTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTG 239

Qy        244 ATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGG 303
              |||||||| ||| ||||| |||||||||||||||| |||||||||  |  ||  || ||
Db        240 ATGTCGACAGTGTCAAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGA 299

Qy        304 CAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCT 363
                 ||||||| || || ||| ||||||||| || ||| ||||| |||||||||| ||||
Db        300 ACAACGTGGAGGACCTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCT 359

Qy        364 CCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTA 423
              |||| || |||||||| |||||||||||| |||||||||||||||||| ||||||||| | |
Db        360 CCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCA 419

Qy        424 TAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGT 483
              |||||||    |||||||| |||||     |||||| |||| ||||| ||||| |||||||
Db        420 TAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGT 479

Qy        484 ACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGA 543
              ||||||| |||||||||||||| || ||||| ||||||| |  |||||||||||| ||||
Db        480 ACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGA 539

Qy        544 TTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAG 603
              | |||||||||||||||||||||||||||| || |||||||||| |||||||||||||
Db        540 TCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAG 599

Qy        604 GC---------------------------------------------------------- 605
              ||
Db        600 GCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGG 659

Qy        606 ----------------------------------------------------------CATAATC 612
                                                                         | ||||
Db        660 ATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATC 719

Qy        613 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 672
              |||||||||| |||||||| |||    |  |||||| |  ||| | || ||||||||||
Db        720 GAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGA 779

Qy        673 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 732
              | || ||||| |  |||||||||||||||||| |||||||  || |||||||| ||||| |
Db        780 ACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTG 839

Qy        733 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 792
              |||||||||||||||||||||||||||| || ||||| ||||||||||||||    ||||
Db        840 AAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCT 899

Qy        793 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 852
              |||||||||||||| |||| |||||||| |||||||| ||||| |||||||||| || |
Db        900 TGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAA 959

Qy        853 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 912
              || ||||||| |||||||||    |||| |||||||| ||||| || ||||||||||||||
Db        960 TACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAG 1019

Qy        913 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 972
              | ||||| |||||||||||||| ||||| |  | || |||||| |||||||| ||||| |
Db       1020 TGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACG 1079

Qy        973 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1032
              ||  |||||||||||  ||||| ||||||||||||||   ||||||||||||||||||||
Db       1080 GCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATT 1139

Qy       1033 TGGGCATCAAATATTCGTT---------------------------------------- 1051
              |||||||||||||||||||
Db       1140 TGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGC 1199
```

```
Qy       1052 ----------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAG 1100
                       ||||||||||||||| |||  |||||||||| |||||  ||||||||||||
Db       1200 TGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAATAG 1259

Qy       1101 CTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTT 1160
              |||||||||||| |||||   |||||||||      | | |       ||     ||||
Db       1260 TTTGGCATGTCATCAGGAACACTTAATGCCC------TAACCTCCGCTCTCATTATTTTT 1313

Qy       1161 AATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT--TATTTTTAAGTTTTATCC 1218
              | |||| ||||| |||||| ||  ||   ||| || || |  ||  || ||    | |||
Db       1314 ATTTTATTGATTTCAGCAACACTTAACTGTTGCATTAGCTTCTAAGTTGAATCAGTTTCC 1373

Qy       1219 GTAGTTTTGATAAAAGAT 1236
              | |||||| | ||  ||
Db       1374 TTGGTTTTGTTGAAGAAT 1391
```

RESULT 14
AF164524
LOCUS AF164524 1490 bp mRNA linear ROD 05-APR-2000
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi) mRNA, complete cds.
ACCESSION AF164524
VERSION AF164524.1 GI:7416966
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1490)
AUTHORS Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
TITLE Characterization of mouse thrombin-activatable fibrinolysis inhibitor
JOURNAL Thromb. Haemost. 83 (2), 297-303 (2000)
MEDLINE 20201996
PUBMED 10739389
REFERENCE 2 (bases 1 to 1490)
AUTHORS Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100, Utrecht 3584 CX, The Netherlands
FEATURES Location/Qualifiers
source 1. .1490
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:477035"
gene 1. .1490
/gene="Tafi"
CDS 84. .1352
/gene="Tafi"
/function="carboxypeptidase"
/codon_start=1
/product="thrombin-activatable fibrinolysis inhibitor"
/protein_id="AAF62385.1"
/db_xref="GI:7416967"
/translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT 438 a 337 c 322 g 393 t
ORIGIN

Query Match 41.7%; Score 656; DB 10; Length 1490;
Best Local Similarity 72.2%; Pred. No. 6.6e-141;
Matches 1009; Conservative 0; Mismatches 215; Indels 174; Gaps 5;

```
Qy          4 AAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCT 63
              || | |||||||||||||||||    |||||| | |||| ||| ||||  | ||||
Db         70 AAGTCACTGTTGGGATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTA-- 127

Qy         64 GTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCT 123
               |||||||||| |||||| || |||||||||| ||| || |||||||||| |||||||
Db        128 -TGAGCAGCATGGCTTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCT 186

Qy        124 CTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGC 183
              | |||||||||||| | || |||||||||||||||||| |||||| | ||||||||||||
Db        187 CCAGGCAAGTTCAACTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGC 246

Qy        184 CGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTG 243
              | || ||||||||  | || | |||||||||  ||||||| |||||||| ||||| ||||
Db        247 CAGTGACAGCTGAATTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTG 306

Qy        244 ATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGG 303
              |||||||| ||| ||||| |||||||||||||| |||||||||  |  ||  || ||
Db        307 ATGTCGACAGTGTCAAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGA 366

Qy        304 CAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCT 363
                |||||| || || ||| ||||||||| || ||| ||||| |||||||||| ||||
Db        367 ACAACGTGGAGGACCTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCT 426

Qy        364 CCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTA 423
              |||| || |||||||| |||||||||||||||||||||||||| |||||||| | |
Db        427 CCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCA 486

Qy        424 TAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGT 483
              |||||||   ||||||||| |||||    |||||| |||| ||||| ||||| |||||||
Db        487 TAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGT 546

Qy        484 ACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGA 543
              |||||| |||||||||||||| || ||||| ||||||| |  |||||||||||| ||||
Db        547 ACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGA 606

Qy        544 TTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAG 603
              | |||||||||||||||||||||||||||| || |||||||||| ||||||||||||
Db        607 TCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAG 666

Qy        604 GC---------------------------------------------------------- 605
              ||
Db        667 GCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGG 726

Qy        606 ----------------------------------------------------------CATAATC 612
                                                                         | ||||
Db        727 ATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATC 786

Qy        613 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 672
              |||||||||| |||||||| |||    |   |||||| |  ||| | || ||||||||||
Db        787 GAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGA 846

Qy        673 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 732
              | || ||||| |  |||||||||||||||||| |||||| || |||||||| ||||| |
Db        847 ACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTG 906

Qy        733 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 792
              |||||||||||||||||||||||||||| || ||||| ||||||||||||||    ||||
Db        907 AAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCT 966

Qy        793 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 852
              |||||||||||||| |||| |||||||| |||||||| ||||| |||||||||| || |
Db        967 TGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAA 1026

Qy        853 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 912
              || ||||||| |||||||||   |||| |||||||| ||||| || ||||||||||||
Db       1027 TACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAG 1086

Qy        913 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 972
              | ||||| |||||||||||||| ||||| |  | || |||||| |||||||| ||||| |
Db       1087 TGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACG 1146

Qy        973 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1032
              ||  |||||||||||  ||||| ||||||||||||||   ||||||||||||||||||||
```

```
Db       1147 GCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATT 1206

Qy       1033 TGGGCATCAAATATTCGTT---------------------------------------- 1051
              |||||||||||||||||||
Db       1207 TGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGC 1266

Qy       1052 -----------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAG 1100
                         ||||||||||||||| ||| |||||||||| |||||  ||||||||||||
Db       1267 TGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAATAG 1326

Qy       1101 CTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTT 1160
              |||||||||||| |||||    |||||||||      | | |      ||      ||||
Db       1327 TTTGGCATGTCATCAGGAACACTTAATGCCC------TAACCTCCGCTCTCATTATTTTT 1380

Qy       1161 AATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT--TATTTTTAAGTTTTATCC 1218
              | |||| ||||| |||||| ||  ||   ||| || || |  ||  || ||    | |||
Db       1381 ATTTTATTGATTTCAGCAACACTTAACTGTTGCATTAGCTTCTAAGTTGAATCAGTTTCC 1440

Qy       1219 GTAGTTTTGATAAAAGAT 1236
                | |||||| | ||  ||
Db       1441 TTGGTTTTGGTGAAGAAT 1458


RESULT 15
AF186188
LOCUS       AF186188               1430 bp    mRNA    linear   ROD 01-OCT-1999
DEFINITION  Mus musculus carboxypeptidase U mRNA, complete cds.
ACCESSION   AF186188
VERSION     AF186188.1  GI:6003651
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Isolation and characterization of mouse liver carboxypeptidase B
            gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Division of Hematology, Washington
            University School of Medicine, 216 S. Kingshighway, St. Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="B6/CBAF1J"
                     /db_xref="taxon:10090"
     CDS             20. .1288
                     /note="pro-carboxypeptidase-B; thrombin-activatable
                     fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
                     plasminogen"
                     /codon_start=1
                     /product="carboxypeptidase U"
                     /protein_id="AAF00528.1"
                     /db_xref="GI:6003652"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVLTEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIGSFNKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      421 a    320 c    310 g    379 t
ORIGIN

  Query Match          41.6%;  Score 654;  DB 10;  Length 1430;
  Best Local Similarity   72.4%;  Pred. No. 1.9e-140;
  Matches  989;  Conservative    0;  Mismatches  205;  Indels  172;  Gaps    4;
```

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||| |  |||||||||||||||||     |||||| | |||| ||| ||||  | ||||
Db          3 AGAAAGTCACTGTTGGGATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
                |||||||||| |||||| || |||||||||| ||| || |||||||||| ||||
Db         63 A---TGAGCAGCATGGCTTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAA 119

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              |||| |||||||||||| | || |||||||||||||||| |||||| | ||||||||||
Db        120 CCTCCAGGCAAGTTCAACTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGC 179

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              |||| || ||||||||  | || | |||||||||  ||||||| |||||||| ||||| |
Db        180 AGCCAGTGACAGCTGAATTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGT 239

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              |||||||||||| ||| ||||| ||||||||||||||| |||||||||  |  ||  ||
Db        240 CTGATGTCGACAGTGTCAAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGA 299

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||    ||||||| || || ||| ||||||||| || ||| ||||| ||||||||||| |
Db        300 TGAACAACGTGGAGGACCTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCG 359

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||| || |||||||| |||||||||| |||||||||||||||||| |||||||||
Db        360 CCTCCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAG 419

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              |  |||||||   |||||||| |||||    |||||| |||| ||||| ||||| ||||
Db        420 TCTTAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGA 479

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              |||||||||| |||||||||||||||| || ||||| ||||||| |  ||||||||||| |
Db        480 AGTACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCT 539

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              |||| |||||||||||||||||||||||||| || |||||||||| ||||||||||
Db        540 GGATCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCA 599

Qy        601 TAGGC---------------------------------------------------------- 605
              |||||
Db        600 TAGGCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACG 659

Qy        606 ----------------------------------------------------------CATA 609
                                                                         | |
Db        660 TGGATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGA 719

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              |||||||||||| |||||||| |||    |  |||||| |  ||| | || |||||||
Db        720 ATCGAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACC 779

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              |||| || ||||| |  ||||||||||||||||| ||||||| || ||||||||  ||||
Db        780 TGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCT 839

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              | |||||||||||||||||||||||||||||||| || ||||| ||||||||||||   |
Db        840 CTGAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACT 899

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              |||||||||||||||| |||| ||||||| |||||||  ||||| |||||||||| |
Db        900 TCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAAC 959

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              | ||| |||||||  ||||||||||   |||| |||||||| ||||| || ||||||||||
Db        960 AAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTC 1019

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||| ||||| ||||||||||||||| |||| |    || |||||| |||||||| ||||
Db       1020 TAGTGGCCAGCGAAGCAGTTCGTGCAATTGGAAGTTTTAATAAAAACACCAGGTACACAC 1079
```

```
Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              | |||  ||||||||||  ||||| ||||||||||||||   ||||||||||||||||
Db       1080 ACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATG 1139

Qy       1030 ATTTGGGCATCAAATATTCGTT---------------------------------------- 1051
              |||||||||||||||||||||
Db       1140 ATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCT 1199

Qy       1052 --------------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                            ||||||||||||||| |||  ||||||||| |||||  |||||||||
Db       1200 TGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1259

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              |||  |||||||||||| |||||   |||||||||      | | |      ||      |
Db       1260 TAGTTTGGCATGTCATCAGGAACACTTAATGCCC------TAACCTCCGCTCTCATTATT 1313

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTAT 1203
              |||| |||| ||||| |||||| ||  ||   ||| || || || |
Db       1314 TTTATTTTATTGATTTCAGCAACACTTAACTGTTGCATTAGCTTCT 1359
```

Search completed: January 1, 2004, 13:55:37
Job time : 5956.84 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:40:19 ; Search time 461.88 Seconds
(without alignments)
9193.336 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1573	100.0	1573	21	AAC81962	Human brain carbox
	2	1469.8	93.4	1625	25	ABX13670	Human protease cDN
	3	1372	87.2	1728	24	ABN95899	Gene #2397 used to
	4	1372	87.2	1749	17	AAT11671	Human plasma carbo
	5	1372	87.2	1749	18	AAT62846	Human plasma carbo
	6	1370.4	87.1	1749	14	AAQ41001	Human plasma carbo
	7	918	58.4	1272	20	AAV74302	Human plasma carbo
	8	853.8	54.3	1400	21	AAF18005	Lung cancer associ
	9	478.6	30.4	55827	25	ABX13671	Human protease gen
	10	331.2	21.1	416	25	ABX49430	Bovine EST associa
	11	324	20.6	333	20	AAH85957	Human single nucle
	12	324	20.6	333	20	AAH87390	Human single nucle
	13	286.8	18.2	397	25	ABX43260	Bovine EST associa
	14	241.6	15.4	251	19	AAX11374	Human biallelic po
	15	167	10.6	231	24	ABN19133	Human ORFX polynuc
c	16	137.2	8.7	431	24	ABK62561	Rat sequence diffe
	17	124	7.9	927	17	AAT35760	Rat mature carboxy
	18	123.4	7.8	921	16	AAQ90601	Porcine carboxypep
	19	123.4	7.8	1215	16	AAQ90600	Porcine Tyr-His-Me
c	20	120	7.6	121	20	AAH85710	Human single nucle
c	21	120	7.6	121	20	AAH85711	Human single nucle
	22	109	6.9	1263	19	AAV41795	Human pancreatic c
	23	107.6	6.8	2023	25	ABZ75116	Anti-human seminal
	24	107.4	6.8	999	17	AAT42500	Mature HCPB coding
	25	107.4	6.8	1053	17	AAT42497	mature HCPB-(His)6
	26	107.4	6.8	1263	17	AAT42494	Human pancreatic c
	27	107.4	6.8	1284	17	AAT42506	ProHCPB gene with
	28	104.2	6.6	1053	20	AAX24804	Human carboxypepta
	29	104.2	6.6	1059	17	AAT42511	Modified HCPB (D25
	30	104.2	6.6	1059	17	AAT42512	Modified HCPB (D25
	31	104.2	6.6	1059	18	AAT62787	Carboxypeptidase B
	32	104.2	6.6	1059	18	AAT62788	Carboxypeptidase B
	33	102.6	6.5	1059	18	AAT62805	Carboxypeptidase B
	34	102.6	6.5	1059	18	AAT62806	Carboxypeptidase B
	35	102.6	6.5	1059	18	AAT62791	Carboxypeptidase B
	36	102.6	6.5	1059	18	AAT62801	Carboxypeptidase B
	37	101	6.4	1059	18	AAT62789	Carboxypeptidase B
	38	101	6.4	1059	18	AAT62790	Carboxypeptidase B
	39	101	6.4	1870	18	AAV17322	DNA encoding human
	40	101	6.4	1870	19	AAV41819	Human carboxypepti
	41	101	6.4	2154	18	AAV17331	PreproHCPB-linker-
	42	99.4	6.3	1059	18	AAT62808	Carboxypeptidase B
	43	99.4	6.3	1059	18	AAT62804	Carboxypeptidase B
	44	99.2	6.3	1059	18	AAT62800	Carboxypeptidase B
	45	97.8	6.2	1059	18	AAT62807	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.
XX
AC AAC81962;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B cDNA.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 18..1100
FT /*tag= a
FT /product= "carboxypeptidase B"
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX

PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
DR P-PSDB; AAB11457.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 2b; Page 64-68; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

Query Match 100.0%; Score 1573; DB 21; Length 1573;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1573; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600

Qy        601 TAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCG 660

Qy        661 GAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTT 720

Qy        721 CCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAG 780

Qy        781 TGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCAT 840

Qy        841 ACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGG 900

Qy        901 AACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCA 960

Qy        961 GGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATT 1020

Qy       1021 GGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTT 1080

Qy       1081 TGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TGCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTA 1140

Qy       1141 TCATTCTGCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TCATTCTGCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT 1200

Qy       1201 TATTTTTAAGTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TATTTTTAAGTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAG 1260

Qy       1261 AGAACCTAATAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 AGAACCTAATAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCT 1320

Qy       1321 TTAAAAAAAAAATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TTAAAAAAAAAATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACT 1380

Qy       1381 AGCCATCTCAAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGAT 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AGCCATCTCAAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGAT 1440

Qy       1441 CATCTCACGCTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTA 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CATCTCACGCTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTA 1500

Qy       1501 AAGATTTCTGCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 AAGATTTCTGCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAA 1560

Qy       1561 AAAAAAAAAAAAA 1573
              |||||||||||||
Db       1561 AAAAAAAAAAAAA 1573
```

RESULT 2
ABX13670
ID ABX13670 standard; cDNA; 1625 BP.
XX
AC ABX13670;
XX
DT 14-FEB-2003 (first entry)
XX

```
DE   Human protease cDNA.
XX
KW   Human; gene; ss; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..16
FT                   /*tag= a
FT   CDS             17..1177
FT                   /*tag= b
FT                   /product= "Protease"
FT   variation       replace(521,A)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR           1178..1625
FT                   /*tag= d
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   P-PSDB; ABG72093, ABG72095.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 4; Fig 1; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   cDNA, the gene for which is located on chromosome 13.
XX
SQ   Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;
```

```
  Query Match             93.4%;  Score 1469.8;  DB 25;  Length 1625;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches 1553;  Conservative    0;  Mismatches   17;  Indels   53;  Gaps    2;

Qy          2 GAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTT 60

Qy         62 CTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAAC 121
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         61 CTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAAC 120

Qy        122 CTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCA 181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCA 180

Qy        182 GCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATC 240

Qy        242 TGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCT 300

Qy        302 GGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGC 361
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGC 360

Qy        362 CTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATT 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATT 420

Qy        422 TATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAA 480

Qy        482 GTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATG 541
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        481 GTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATG 540

Qy        542 GATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCAT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCAT 600

Qy        602 AGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGG 660

Qy        662 AACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTC 721
              |||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||
Db        661 AACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTC 720

Qy        722 CTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGT 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGT 780

Qy        782 GGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATA 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATA 840

Qy        842 CTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGA 900

Qy        902 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAG 961
              |||||||||||||||||||||||||||||||||||||||| |||| ||||||||||||||
Db        901 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAG 960

Qy        962 GTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTG 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTG 1020
```

```
Qy       1022 GATCTATGATTTGGGCATCAAATATTCGT------------------------------ 1050
              ||||||||||||||||||||||||||||
Db       1021 GATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATA 1080

Qy       1051 ---------------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGT 1089
                                   ||||||||||||||||||||||||||||||||||||||
Db       1081 CGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGT 1140

Qy       1090 CTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGC 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGC 1200

Qy       1150 TTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAA 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAA 1260

Qy       1210 GTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAA 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAA 1320

Qy       1270 TAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAA 1329
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1321 TAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAA 1379

Qy       1330 AATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTC 1389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 AATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTC 1439

Qy       1390 AAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACG 1449
              |||||| |||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1440 AAGCAAGTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAATGATCATCTCACG 1499

Qy       1450 CTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCT 1509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 CTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCT 1559

Qy       1510 GCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAAAAAAAA 1569
              |||||||||||||||||||||||| ||| | | || |||||||||| ||||||||||||
Db       1560 GCTCCAAATTTTCAATAAATTTCTGCTTGTGCCTTTAAAAAAAAAAATAAAAAAAAAAAA 1619

Qy       1570 AAA 1572
              | |
Db       1620 ATA 1622
```

```
RESULT 3
ABN95899
ID   ABN95899 standard; DNA; 1728 BP.
XX
AC   ABN95899;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2397 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 2397; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;
```

Query Match 87.2%; Score 1372; DB 24; Length 1728;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 122

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 542

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 602

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
```

```
Db        603 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 662

Qy        609 ------------------------------------------------------------A 609
                                                                           |
Db        663 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 722

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        723 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 782

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        783 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 842

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        843 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 902

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 962

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        963 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1022

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1023 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1082

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1142

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db       1143 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1202

Qy       1051 -------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1203 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1263 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1322

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1323 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1382

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1383 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1442

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1443 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1502

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1503 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1562

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1563 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1622

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1623 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1682

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
              |||||||||||||||||||||||||||||||||||||||||||||
Db       1683 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1728
```

```
RESULT 4
AAT11671
ID   AAT11671 standard; DNA; 1749 BP.
XX
AC   AAT11671;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human plasma carboxypeptidase B.
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   P-PSDB; AAR90293.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

  Query Match            87.2%;  Score 1372;  DB 17;  Length 1749;
  Best Local Similarity   90.3%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
```

```
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              |||||||  |||||||   ||||||||||||||||||||||||||||||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749
```

RESULT 5
AAT62846

```
ID   AAT62846 standard; DNA; 1749 BP.
XX
AC   AAT62846;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human PCPB
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
FT   misc_binding    134..177
FT                   /*tag=  d
FT                   /bound_moeity= 46_bp_probe
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
```

```
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   P-PSDB; AAW14733.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence encodes human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

  Query Match             87.2%;  Score 1372;  DB 18;  Length 1749;
  Best Local Similarity   90.3%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||

Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ------------------------------------------------------------A 609
                                                                           |
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 -------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749
```

```
RESULT 6
AAQ41001
ID   AAQ41001 standard; cDNA; 1749 BP.
XX
AC   AAQ41001;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B gene.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     41..106
FT                   /*tag=  a
FT   mat_peptide     107..1312
FT                   /*tag=  b
FT                   /product= PCPB
FT   misc_feature    134..177
FT                   /*tag=  c
FT                   /function= probe
FT                   /note= "used to obtain full-length clones"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   P-PSDB; AAR36273.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Disclosure; Fig 4; 40pp; English.
XX
CC   Human plasma carboxypeptidase B was isolated from human plasma and
CC   partially sequenced. Oligonucleotide primers were designed based on
CC   the partial amino acid sequences. The primers were used in a PCR
CC   amplification to identify cDNA encoding PCPB from a human liver cDNA
CC   library. The PCR product was capable of encoding the first 37 amino
CC   acids of PCPB; a 46mer probe was used to obtain the full-length
CC   sequence which, although disclosed in the specification, is not
CC   claimed. PCPB inhibits the enzymatic conversion by tPA of
CC   plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

  Query Match           87.1%;  Score 1370.4;  DB 14;  Length 1749;
  Best Local Similarity   90.2%;  Pred. No. 0;
  Matches 1557;  Conservative    0;  Mismatches    6;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||| ||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749
```

RESULT 7
AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX
AC AAV74302;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease; ds.
XX
OS Homo sapiens.
XX
PN WO9855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI; 1999-045800/04.
DR P-PSDB; AAW92270.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Example 1; Page 24; 35pp; English.
XX
CC This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Query Match 58.4%; Score 918; DB 20; Length 1272;
Best Local Similarity 86.8%; Pred. No. 2.8e-211;
Matches 1104; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

```
Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        609 ------------------------------------------------------------ 608

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                     ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1047 TCGT----------------------------------------------------TTAC 1054
              ||||                                                    ||||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy       1115 AGGAATGTTTAA 1126
              ||||||||||||
Db       1261 AGGAATGTTTAA 1272
```

RESULT 8
AAF18005
ID AAF18005 standard; DNA; 1400 BP.
XX
AC AAF18005;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
DR P-PSDB; AAB58129.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 1; Page 507; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

Query Match 54.3%; Score 853.8; DB 21; Length 1400;
Best Local Similarity 93.3%; Pred. No. 8.5e-196;
Matches 934; Conservative 1; Mismatches 13; Indels 53; Gaps 2;

```
Qy        607 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 666
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 AGAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 236

Qy        667 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 726
              |||||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||
Db        237 ACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 296

Qy        727 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 356

Qy        787 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 416

Qy        847 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 476

Qy        907 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATA 966
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        477 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATA 536

Qy        967 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 596

Qy       1027 ATGATTTGGGCATCAAATATTCGT---------------------------------- 1050
              |||||||||||||||||||||||
Db        597 ATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGAT 656

Qy       1051 ----------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTA 1094
                             ||||||||||||||||||||||||||||||||||||||||||||
Db        657 TCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTA 716

Qy       1095 AAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCG 1154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 AAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCG 776

Qy       1155 TATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTT 1214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 TATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTT 836

Qy       1215 ATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGT 1274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        837 ATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGT 896

Qy       1275 GCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTG 1334
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        897 GCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAAAATTG 955

Qy       1335 TAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCA 1394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 TAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCA 1015

Qy       1395 ACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGAT 1454
             ||||||||||||||||||||||||||||||||||||:||||||||||||||||||||
Db       1016 ACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAWGATCATCTCACGCTGAT 1075

Qy       1455 CATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCC 1514
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCC 1135

Qy       1515 AAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAA 1555
              |||||||||||||||||| ||| | | || | ||| | ||
Db       1136 AAATTTTCAATAAATTTCTGCTTGTGCCTTTAGAAATACAA 1176


RESULT 9
ABX13671
ID   ABX13671 standard; DNA; 55827 BP.
XX
AC   ABX13671;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease gene.
XX
KW   Human; gene; ds; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2017..53409
FT                   /*tag= a
FT                   /product= "Protease"
FT   exon            2017..2090
FT                   /*tag= b
FT                   /number= 1
FT   intron          2091..19206
FT                   /*tag= c
FT                   /number= 1
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            19207..19282
FT                   /*tag= d
FT                   /number= 2
FT   intron          19283..22683
FT                   /*tag= e
FT                   /number= 2
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            22684..22808
FT                   /*tag= f
FT                   /number= 3
FT   intron          22809..24477
FT                   /*tag= g
FT                   /number= 3
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            24478..24586
FT                   /*tag= h
FT                   /number= 4
FT   intron          24587..28129
FT                   /*tag= i
FT                   /number= 4
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            28130..28231
FT                   /*tag= j
FT                   /number= 5
FT   intron          28232..33052
FT                   /*tag= k
FT                   /number= 5
FT   exon            33053..33157
FT                   /*tag= l
FT                   /number= 6
FT   intron          33158..42288
FT                   /*tag= m
FT                   /number= 6
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            42289..42382
FT                   /*tag= n
FT                   /number= 7
FT   intron          42383..48643
FT                   /*tag= o
FT                   /number= 7
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            48644..48846
FT                   /*tag= p
FT                   /number= 8
FT   intron          48847..51176
FT                   /*tag= q
FT                   /number= 8
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            51177..51264
FT                   /*tag= r
FT                   /number= 9
FT   intron          51265..53224
FT                   /*tag= s
FT                   /number= 9
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            53225..53409
FT                   /*tag= t
FT                   /number= 10
FT   variation       replace(858,T)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(2122,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4088,A)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4240,A)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4416..4418,TA)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4434,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4969,C)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5868,C)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6254,C)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10171,C)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10452,A)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11613,A)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
```

```
FT   variation       replace(12130,T)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17867..17868,GAA)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18243,G)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22450,C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23003,T)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24055..24056,GAA)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24132,C)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25713,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28476,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29404,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31047,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31445,C)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31447,C)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31600,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31714,C)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31715,T)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32193,G)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32341,G,A)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32561,G)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32600..35601,AAA)
FT                   /*tag= az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32642,T)
FT                   /*tag= ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32793..32794,TAG)
FT                   /*tag= bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33071,A)
FT                   /*tag= bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(34721,T)
FT                   /*tag= bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35304,A)
FT                   /*tag= be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35425,C)
FT                   /*tag= bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(36050,G)

  Query Match            30.4%;  Score 478.6;  DB 25;  Length 55827;
  Best Local Similarity  98.0%;  Pred. No. 7.6e-105;
  Matches  495;  Conservative    0;  Mismatches    9;  Indels    1;  Gaps    1;

Qy       1051 TTACATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAATAGCTTGGCATGT 1110
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53334 TTACATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAATAGCTTGGCATGT 53393

Qy       1111 CATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTTAATTTACTGA 1170
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53394 CATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTTAATTTACTGA 53453

Qy       1171 TTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCCGTAGTTTTGATA 1230
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53454 TTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCCGTAGTTTTGATA 53513

Qy       1231 AAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTACTTTGCCATTAA 1290
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53514 AAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTACTTTGCCATTAA 53573

Qy       1291 GGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAAAAGTCTAGTTACC 1350
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      53574 GGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAAAATTGTAAAAGTCTAGTTACC 53632

Qy       1351 TACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTTTCGACGTTTGAC 1410
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53633 TACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTTTCGACGTTTGAC 53692

Qy       1411 TAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCATTGGATCCTACTCA 1470
             |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      53693 TAGCCATCTCAAGCAAGTTTAATCAATGATCATCTCACGCTGATCATTGGATCCTACTCA 53752

Qy       1471 ACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAATTTTCAATAAATT 1530
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53753 ACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAATTTTCAATAAATT 53812

Qy       1531 TCTTCTTCTCCTTTAAAAAAAAAAA 1555
             ||| ||| | | || | ||| | ||
Db      53813 TCTGCTTGTGCCTTTAGAAATACAA 53837


RESULT 10
ABX49430
ID   ABX49430 standard; cDNA; 416 BP.
XX
AC   ABX49430;
XX
DT   21-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #14595.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;   99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
```

PA (BYAT/) BYATT J C.
PA (MATH/) MATHIALAGAN N.
PA (TAON/) TAO N.
PA (WARR/) WARREN W C.
XX
PI Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR WPI; 2003-110599/10.
XX
PT New nucleic acid associated with lactation, and muscle and fat
PT deposition, useful for genome mapping, gene identification and
PT analysis, cattle breeding, or for genetically improving cattle -
XX
PS Claim 2; SEQ ID No 14595; 245pp; English.
XX
CC The invention relates to a purified nucleic acid molecule associated with
CC lactation or muscle and fat deposition (designated LMFD), derived
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC second nucleic acid molecule comprising any of 15112 nucleotide
CC sequences, appearing as ABX34836-ABX49947, or complements of them.
CC Also included are; (1) a transformed cell having a nucleic acid
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC translated sequence that functions in the cell to cause termination of
CC transcription and addition of polyadenylated ribonucleotides to a 3' end
CC of the mRNA molecule; and (2) determining a level or pattern of a
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC complement or fragment) with a complementary nucleic acid molecule
CC obtained from the bovine cell or tissue, where hybridisation between the
CC marker nucleic acid and the complementary nucleic acid permits the
CC detection of the molecule; and (b) detecting the level or pattern of the
CC complementary nucleic acid, where the detection of the complementary
CC nucleic acid is predictive of the level or pattern of the molecule.
CC The LMFD nucleic acid is used for determining a level or pattern
CC of a molecule in a bovine cell or tissue. It is useful for genome
CC mapping, gene identification and analysis, cattle breeding, preparation
CC of constructs for use in cattle gene expression, or for genetically
CC improving cattle. The present sequence is one of the 15112 bovine
CC LMFD EST (expressed sequence tag) nucleic acids.
CC Note: The present sequence was not shown in the specification but
CC was obtained in electronic format from the USPTO web site:
CC seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ Sequence 416 BP; 125 A; 86 C; 93 G; 112 T; 0 other;

Query Match 21.1%; Score 331.2; DB 25; Length 416;
Best Local Similarity 87.3%; Pred. No. 5e-70;
Matches 363; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy      167 GATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTT 226
            |||||||||||||||||| ||| |||||||   |||||||||  | |  ||||||||||
Db        1 GATTGTTCTCTGGCAGCCAGTAGCAGCTGAATATATTGTGAAGGGATACGAAGTCCATTT 60

Qy      227 TTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCC 286
            ||||||||||||||||||   |||||||||||||||||||||||| |||| |||||||
Db       61 TTTTGTGAATGCATCTGATGTAAGCAATGTGAAAGCCCATTTAAATGCGAGCAGAATTCC 120

Qy      287 ATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACAC 346
            || ||| ||| || ||| | | ||||||||||||||| |  ||||||| |||||| |||||
Db      121 ATTCAGGGTCCTGGTGGAAAATGTGGAAGATCTTATCCGGCAGCAGACTTCCAATGACAC 180

Qy      347 AGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTA 406
              |||||||||| || ||| | || |||||||||||||||||||||||||||||| |||||
Db      181 CATCAGCCCCCGGGCATCCTCCTCCTACTATGAACAGTATCACTCACTAAATGAGATCTA 240

Qy      407 TTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGG 466
            |||||||||||| |||| |||||| || |||||||||| ||  ||||||||||||||||
Db      241 TTCTTGGATAGAAGTTATGACTGAGCGGTATCCTGATATGGTTGAAAAAATCCACATTGG 300

Qy      467 ATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGC 526
            ||||||||  |||||||||||||| |||||||||||||||||  ||||||||||   ||
Db      301 ATCCTCATACGAGAAGTACCCACTTTATGTTTTAAAGGTTTCTAAAAAAGAACAAAGGGC 360
```

Qy 527 CAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTG 582
||||||||||| |||||||||||||||||||| |||||||| ||||||||||||
Db 361 CAAAAATGCCATGTGGATTGACTGTGGAATCCACGCCAGAGAGTGGATCTCTCCTG 416

RESULT 11
AAH85957
ID AAH85957 standard; DNA; 333 BP.
XX
AC AAH85957;
XX
DT 27-FEB-2002 (first entry)
XX
DE Human single nucleotide polymorphism containing DNA sequence #814.
XX
KW Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
KW transgenic; single nucleotide polymorphism; SNP; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Variation replace(274,C,G)
FT /*tag= a
FT /standard_name= "single nucleotide polymorphism"
XX
PN WO9953095-A2.
XX
PD 21-OCT-1999.
XX
PF 30-MAR-1999; 99WO-US06893.
XX
PR 09-APR-1998; 98US-0057871.
XX
PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
PI Lander ES, Wang D, Hudson T;
XX
DR WPI; 1999-620443/53.
XX
PT Polymorphic human genomic sequences and related allele-specific probes
PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring
PT of disease -
XX
PS Claim 1; Page 117; 330pp; English.
XX
CC This invention describes novel human nucleic acid segments (I)
CC containing polymorphic sites. The polynucleotides of (I) are used for,
CC e.g. correlating disease polymorphisms (or disease susceptibility) or
CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC response to drugs etc.); diagnosing and monitoring e.g. cancer,
CC inflammation, heart or central nervous system diseases; detecting
CC susceptibility to microbial infection; treating or preventing such
CC diseases; forensic analysis; gene therapy; paternity testing; mapping
CC genomic loci associated with phenotypic traits (and subsequent cloning
CC of the genes responsible); and the production of transgenic organisms.
CC Antibodies raised against (I) are useful as diagnostic and therapeutic
CC tools and in drug screening. AAH85144 - AAH87644 represent the human
CC DNA sequences containing biallelic polymorphic sites described in the
CC invention.
XX
SQ Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other;

Query Match 20.6%; Score 324; DB 20; Length 333;
Best Local Similarity 97.3%; Pred. No. 2.5e-68;
Matches 324; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```
Qy     1159 TTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCC 1218
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 TTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCC 60

Qy     1219 GTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTA 1278
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         61 GTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTA 120

Qy       1279 CTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAAA 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||        |||||||
Db        121 CTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTNNNNNNNNNTTGTAAA 180

Qy       1339 AGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTT 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTT 240

Qy       1399 TCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCATT 1458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCATT 300

Qy       1459 GGATCCTACTCAACAAAAGGAAGGGTGGTCAGA 1491
              |||||||||||||||||||||||||||||||
Db        301 GGATCCTACTCAACAAAAGGAAGGGTGGTCAGA 333


RESULT 12
AAH87390
ID   AAH87390 standard; DNA; 333 BP.
XX
AC   AAH87390;
XX
DT   27-FEB-2002  (first entry)
XX
DE   Human single nucleotide polymorphism containing DNA sequence #2247.
XX
KW   Biallelic marker; polymorphism; human; disease; diagnosis; treatment;
KW   phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;
KW   transgenic; single nucleotide polymorphism; SNP; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Variation       replace(238,G)
FT                   /*tag=  a
FT                   /standard_name= "single nucleotide polymorphism"
XX
PN   WO9953095-A2.
XX
PD   21-OCT-1999.
XX
PF   30-MAR-1999;   99WO-US06893.
XX
PR   09-APR-1998;   98US-0057871.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Lander ES,  Wang D,  Hudson T;
XX
DR   WPI; 1999-620443/53.
XX
PT   Polymorphic human genomic sequences and related allele-specific probes
PT   and primers, useful for genetic analysis, e.g. diagnosis and monitoring
PT   of disease -
XX
PS   Claim 1; Page 283; 330pp; English.
XX
CC   This invention describes novel human nucleic acid segments (I)
CC   containing polymorphic sites. The polynucleotides of (I) are used for,
CC   e.g. correlating disease polymorphisms (or disease susceptibility) or
CC   other phenotypic traits (e.g. baldness, obesity, fertility, strength,
CC   response to drugs etc.); diagnosing and monitoring e.g. cancer,
CC   inflammation, heart or central nervous system diseases; detecting
CC   susceptibility to microbial infection; treating or preventing such
CC   diseases; forensic analysis; gene therapy; paternity testing; mapping
CC   genomic loci associated with phenotypic traits (and subsequent cloning
CC   of the genes responsible); and the production of transgenic organisms.
CC   Antibodies raised against (I) are useful as diagnostic and therapeutic
CC   tools and in drug screening. AAH85144 - AAH87644 represent the human
CC   DNA sequences containing biallelic polymorphic sites described in the
CC   invention.
XX
SQ   Sequence 333 BP; 90 A; 64 C; 53 G; 117 T; 9 other;

  Query Match             20.6%;  Score 324;  DB 20;  Length 333;
  Best Local Similarity   97.3%;  Pred. No. 2.5e-68;
  Matches  324;  Conservative    0;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1159 TTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCC 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCC 60

Qy       1219 GTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTA 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTA 120

Qy       1279 CTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAAA 1338
              ||||||||||||||||||||||||||||||||||||||||||||||||        |||||||
Db        121 CTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTNNNNNNNNNTTGTAAA 180

Qy       1339 AGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTT 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTT 240

Qy       1399 TCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCATT 1458
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCATT 300

Qy       1459 GGATCCTACTCAACAAAAGGAAGGGTGGTCAGA 1491
              |||||||||||||||||||||||||||||||
Db        301 GGATCCTACTCAACAAAAGGAAGGGTGGTCAGA 333


RESULT 13
ABX43260
ID   ABX43260 standard; cDNA; 397 BP.
XX
AC   ABX43260;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #8425.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;   99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle -
XX
PS   Claim 2; SEQ ID No 8425; 245pp; English.
XX
```

```
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ   Sequence 397 BP; 100 A; 94 C; 96 G; 107 T; 0 other;

  Query Match             18.2%;  Score 286.8;  DB 25;  Length 397;
  Best Local Similarity   83.9%;  Pred. No. 2.4e-59;
  Matches  324;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy         12 GTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAG 71
              || |||||||||||||  ||| | | |||||||||  |||  || || ||||||| | ||
Db         12 GTGGGGATGAAGCTTTATAGCTTGGGAGTCCTTGTCGCCACCGTCCTGTTCTGTGGGGAG 71

Qy         72 CATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAA 131
              |||| |||||| || ||||| ||||| ||| || ||||||||||||||||||| ||||||
Db         72 CATGCCTTCGCCTTTCAGAGGGGCCAGGTTTTATCTGCTCTTCCTAGAACCTCCAGGCAA 131

Qy        132 GTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACA 191
              |||||| |||| |||||| |||||||||||||  ||||||||||||||||||| ||| ||
Db        132 GTTCAAATTCTGCAGAATGTTACTACAACATACAAGATTGTTCTCTGGCAGCCAGTAGCA 191

Qy        192 GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGAC 251
              |||||   ||||||||||  | |  |||||||||||||||| ||||||||||||||   |
Db        192 GCTGAATATATTGTGAAGGGATACGAAGTCCATTTTTTTTGTGAATGCATCTGATGTAAGC 251

Qy        252 AATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTG 311
              |||||||||||||||||||||| |||| ||||||||| ||| ||| || ||| | | |||
Db        252 AATGTGAAAGCCCATTTAAATGCGAGCAGAATTCCATTCAGGGTCCTGGTGGAAAATGTG 311

Qy        312 GAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCG 371
              |||||||||| |  ||||||| |||||| |||||  |||||||||| || ||| | ||
Db        312 GAAGATCTTATCCGGCAGCAGACTTCCAATGACACCATCAGCCCCCGGGCATCCTCCTCC 371

Qy        372 TACTATGAACAGTATCACTCACTAAA 397
              |||||||||||||||||||||||||
Db        372 TACTATGAACAGTATCACTCACTAAA 397


RESULT 14
AAX11374
ID   AAX11374 standard; DNA; 251 BP.
XX
AC   AAX11374;
XX
DT   30-MAR-1999  (first entry)
XX
DE   Human biallelic polymorphic DNA fragment WI-7773b.
XX
KW   Polymorphism; biallelic; human; forensic; paternity testing; disease;
KW   detection; phenotypic typing; characteristic; infection; hereditary;
KW   autoimmune disease; cancer; inflammation; drug; therapy; medicament;
KW   treatment; marker; ss.
XX
OS   Homo sapiens.
XX
PN   WO9820165-A2.
XX
PD   14-MAY-1998.
XX
PF   05-NOV-1997;   97WO-US20313.
XX
PR   06-NOV-1996;   96US-0030455.
XX
PA   (WHED ) WHITEHEAD INST BIOMEDICAL RES.
XX
PI   Hudson T,  Lander ES,  Wang D;
XX
DR   WPI; 1998-286974/25.
XX
PT   New isolated nucleic acid segments from the human genome - used for
PT   determining polymorphic forms for use in e.g. forensics, paternity
PT   testing or phenotypic typing for disease
XX
PS   Claim 1; Page 152; 310pp; English.
XX
CC   AAX10269-X12937 are human DNA fragments which contain biallelic
CC   polymorphic markers which have been isolated using the primers
CC   represented in AAX09121-X10268. The base occupying the polymorphic site
CC   is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
CC   can be used in methods for determining polymorphic forms in an individual
CC   for use in e.g. forensics, paternity testing or for phenotypic typing for
CC   diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan
CC   syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
CC   familial hypercholesterolemia, polycystic kidney disease, hereditary
CC   spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary
CC   haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
CC   syndrome, osteogenesis imperfecta, acute intermittent porphyria,
CC   autoimmune diseases, inflammation, cancer, diseases of the nervous
CC   system, infection by pathogenic microorganisms, and characteristics such
CC   as longevity, appearance (e.g. baldness, obesity), strength, speed,
CC   endurance, fertility, and susceptibility or receptivity to particular
CC   drugs or therapeutic treatments. The isolated polymorphic nucleic acid
CC   segments can also be used to produce medicaments for the treatment or
CC   prophylaxis of such diseases.
XX
SQ   Sequence 251 BP; 62 A; 44 C; 37 G; 98 T; 10 other;

  Query Match             15.4%;  Score 241.6;  DB 19;  Length 251;
  Best Local Similarity   96.0%;  Pred. No. 1.6e-48;
  Matches  241;  Conservative    1;  Mismatches    9;  Indels    0;  Gaps    0;

Qy       1159 TTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCC 1218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCC 60

Qy       1219 GTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTA 1278
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTA 120

Qy       1279 CTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAAA 1338
              |||||||||||||||||||||||||||||||||||||||||||||          |||||||
Db        121 CTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTNNNNNNNNNTTGTAAA 180

Qy       1339 AGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTT 1398
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db        181 AGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAASTT 240

Qy       1399 TCGACGTTTGA 1409
              ||||||||||||
Db        241 TCGACGTTTGA 251
```

RESULT 15
ABN19133
ID ABN19133 standard; cDNA; 231 BP.
XX
AC ABN19133;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX polynucleotide sequence SEQ ID NO:6743.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
DR WPI; 2002-106308/14.
DR P-PSDB; ABP03381.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure; SEQ ID 6743; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 231 BP; 74 A; 58 C; 40 G; 59 T; 0 other;

Query Match 10.6%; Score 167; DB 24; Length 231;
Best Local Similarity 82.7%; Pred. No. 1.5e-30;
Matches 191; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy      240 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 299
            ||||||||| ||| ||| |||||| |||||||||| |||| |||||||||  |  ||| ||
Db        1 TCTGATGTCAACAGTGTCAAAGCCTATTTAAATGCGAGCAGAATTCCATTTAACGTCCTG 60

Qy      300 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 359
              ||    ||||||| ||||| |||||||||||||  ||||| ||||| || |||||||||
Db       61 ATGAACAACGTGGAGGATCTAATTCAACAGCAGACGTCCAATGACACTGTTAGCCCCCGA 120

Qy      360 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 419
            |||||| | || ||||||||||||||||||||  |||||||||||||||| |||||||||
Db      121 GCCTCCTCCTCATACTATGAACAGTATCACTCGTTAAATGAAATCTATTCCTGGATAGAA 180

Qy      420 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 470
            ||||||||||   ||| ||||| |||||    |||||| ||||||||||
Db      181 GTTATAACTGAACAGCACCCTGACATGCTCCAGAAAATCTACATTGGATCC 231
```

Search completed: January 1, 2004, 11:08:37
Job time : 479.88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 10:51:23 ; Search time 104.353 Seconds
(without alignments)
6653.349 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1469.8	93.4	1625	4	US-09-813-133A-1	Sequence 1, Appli
2	1372	87.2	1749	1	US-07-649-591B-2	Sequence 2, Appli
3	1372	87.2	1749	1	US-08-277-540-2	Sequence 2, Appli
4	1372	87.2	1749	1	US-08-430-787A-2	Sequence 2, Appli
5	918	58.4	1272	2	US-08-869-057-1	Sequence 1, Appli
6	478.6	30.4	55827	4	US-09-813-133A-3	Sequence 3, Appli
7	124	7.9	927	2	US-08-782-760-5	Sequence 5, Appli
8	124	7.9	927	5	PCT-US96-00995-5	Sequence 5, Appli
9	123.4	7.8	921	1	US-08-696-139-3	Sequence 3, Appli
10	123.4	7.8	1215	1	US-08-696-139-1	Sequence 1, Appli
11	109	6.9	1263	4	US-09-011-769A-38	Sequence 38, Appl
12	107.4	6.8	999	2	US-08-860-882A-67	Sequence 67, Appl
13	107.4	6.8	999	4	US-09-011-769A-50	Sequence 50, Appl
14	107.4	6.8	1053	2	US-08-860-882A-64	Sequence 64, Appl
15	107.4	6.8	1053	4	US-09-011-769A-46	Sequence 46, Appl
16	107.4	6.8	1263	2	US-08-860-882A-56	Sequence 56, Appl
17	107.4	6.8	1284	2	US-08-860-882A-71	Sequence 71, Appl
18	107.4	6.8	1284	4	US-09-011-769A-55	Sequence 55, Appl
19	104.2	6.6	1053	4	US-09-463-451-27	Sequence 27, Appl
C 20	104.2	6.6	1053	4	US-09-463-451-28	Sequence 28, Appl
21	104.2	6.6	1059	2	US-08-860-882A-74	Sequence 74, Appl
22	104.2	6.6	1059	2	US-08-860-882A-77	Sequence 77, Appl
23	104.2	6.6	1059	4	US-09-011-769A-59	Sequence 59, Appl
24	104.2	6.6	1059	4	US-09-011-769A-63	Sequence 63, Appl
25	101	6.4	1870	3	US-09-171-945-112	Sequence 112, App
26	101	6.4	2154	3	US-09-171-945-124	Sequence 124, App
27	92.8	5.9	1311	4	US-09-675-305-9	Sequence 9, Appli
28	92.8	5.9	2128	4	US-09-675-305-13	Sequence 13, Appl
29	88.6	5.6	1050	4	US-09-675-305-11	Sequence 11, Appl
30	87.6	5.6	1200	4	US-09-710-099-7	Sequence 7, Appli
31	83.6	5.3	945	4	US-09-710-099-3	Sequence 3, Appli
32	83.6	5.3	945	4	US-09-710-099-11	Sequence 11, Appl
33	78.4	5.0	741	4	US-09-675-305-5	Sequence 5, Appli
34	71.6	4.6	1251	3	US-08-640-906-3	Sequence 3, Appli
35	71.6	4.6	1251	4	US-09-395-936-3	Sequence 3, Appli
36	70.6	4.5	1257	3	US-08-640-906-1	Sequence 1, Appli
37	70.6	4.5	1257	4	US-09-395-936-1	Sequence 1, Appli
38	51.4	3.3	1056	4	US-09-710-099-1	Sequence 1, Appli
39	51.4	3.3	1056	4	US-09-710-099-9	Sequence 9, Appli
40	51.4	3.3	1311	4	US-09-710-099-5	Sequence 5, Appli
41	51.4	3.3	2247	4	US-09-710-099-15	Sequence 15, Appl
42	48	3.1	629	4	US-09-280-116-228	Sequence 228, App
43	46.8	3.0	285	2	US-08-782-760-2	Sequence 2, Appli
44	46.8	3.0	285	5	PCT-US96-00995-2	Sequence 2, Appli
45	45.6	2.9	8920	2	US-08-446-855A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Query Match 93.4%; Score 1469.8; DB 4; Length 1625;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1553; Conservative 0; Mismatches 17; Indels 53; Gaps 2;

```
Qy          2 GAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTT 61
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTT 60

Qy         62 CTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAAC 121
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         61 CTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAAC 120

Qy        122 CTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCA 181
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCA 180

Qy        182 GCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATC 241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATC 240

Qy        242 TGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCT 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCT 300

Qy        302 GGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGC 361
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGC 360

Qy        362 CTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATT 421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        361 CTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATT 420

Qy        422 TATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAA 481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAA 480

Qy        482 GTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATG 541
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        481 GTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATG 540

Qy        542 GATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCAT 601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCAT 600

Qy        602 AGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGG 661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGG 660

Qy        662 AACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTC 721
              ||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db        661 AACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTC 720

Qy        722 CTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGT 781
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGT 780

Qy        782 GGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATA 841
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATA 840

Qy        842 CTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGA 901
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGA 900

Qy        902 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAG 961
              |||||||||||||||||||||||||||||||||||| |||| ||||||||||||||||
Db        901 ACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAG 960

Qy        962 GTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTG 1021
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTG 1020

Qy       1022 GATCTATGATTTGGGCATCAAATATTCGT------------------------------ 1050
              |||||||||||||||||||||||||||
Db       1021 GATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATA 1080

Qy       1051 ---------------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGT 1089
                                   ||||||||||||||||||||||||||||||||||||||
Db       1081 CGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGT 1140

Qy       1090 CTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGC 1149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGC 1200

Qy       1150 TTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAA 1209
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAA 1260

Qy       1210 GTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAA 1269
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAA 1320

Qy       1270 TAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAA 1329
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       1321 TAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAA 1379

Qy       1330 AATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTC 1389
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 AATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTC 1439

Qy       1390 AAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACG 1449
              |||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db       1440 AAGCAAGTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAATGATCATCTCACG 1499

Qy       1450 CTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCT 1509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 CTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCT 1559

Qy       1510 GCTCCAAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAAAAAAA 1569
              |||||||||||||||||||||||| ||| | | || |||||||||| ||||||||||||
Db       1560 GCTCCAAATTTTCAATAAATTTCTGCTTGTGCCTTTAAAAAAAAAAATAAAAAAAAAAAA 1619

Qy       1570 AAA 1572
              |||
Db       1620 ATA 1622
```

RESULT 2
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/649,591B
; FILING DATE: 19910201
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1749 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: hybridization probe
; LOCATION: 133 to 178
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: potential clip site
; LOCATION: 380 to 382
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; FEATURE:
; NAME/KEY: signal sequence
; LOCATION: 41 to 106
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
US-07-649-591B-2

Query Match 87.2%; Score 1372; DB 1; Length 1749;

```
  Best Local Similarity   90.3%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ----------------------------------------------------------A 609
                                                                          |
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||

Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749
```

```
RESULT 3
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
```

```
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-277-540-2

  Query Match             87.2%;  Score 1372;  DB 1;  Length 1749;
  Best Local Similarity   90.3%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ----------------------------------------------------------A 609
                                                                         |
```

```
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749

RESULT 4
US-08-430-787A-2
```

```
; Sequence 2, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-787A-2

  Query Match             87.2%;  Score 1372;  DB 1;  Length 1749;
  Best Local Similarity   90.3%;  Pred. No. 0;
  Matches 1558;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         24 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 83

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         84 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 143

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 203

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 263

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 323

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 383

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 443

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 503

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 563

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 623

Qy        601 TAGGCCAT---------------------------------------------------- 608
              ||||||||
Db        624 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 683

Qy        609 ----------------------------------------------------------A 609
                                                                         |
Db        684 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 743

Qy        610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 803

Qy        670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
              ||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        804 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 863

Qy        730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 923

Qy        790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        924 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 983

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1043

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       1044 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1103

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1104 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1163

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db       1164 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1223

Qy       1051 -------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                          ||||||||||||||||||||||||||||||||||||||||||||||||
Db       1224 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1284 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1343

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1344 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1403

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1404 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1463

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1464 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1523

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1524 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1583

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1584 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1643

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1644 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1703

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1704 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1749


RESULT 5
US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1272 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    PUBLICATION INFORMATION:
;      AUTHORS:  Eaton, Dan L.
;      AUTHORS:  Malloy, Beth E.
;      AUTHORS:  Tsai, Siao P
;      AUTHORS:  Henzel, William
;      AUTHORS:  Drayna, Dennis
;      TITLE:  Isolation, Molecular Cloning, and Partial
;      TITLE:  Characterization of a No. 5985562el Carboxypeptidase B
;      TITLE:  from Human Plasma
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  266
;      ISSUE:  32
;      PAGES:  21833-21838
;      DATE:  No. 5985562 15-1991
US-08-869-057-1

  Query Match            58.4%;  Score 918;  DB 2;  Length 1272;
  Best Local Similarity  86.8%;  Pred. No. 5.6e-213;
  Matches 1104;  Conservative    0;  Mismatches    5;  Indels  163;  Gaps    2;

Qy         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        609 ------------------------------------------------------------ 608

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                       ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
              |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1047 TCGT----------------------------------------------------TTAC 1054
              ||||                                                    ||||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy       1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy       1115 AGGAATGTTTAA 1126
              ||||||||||||
Db       1261 AGGAATGTTTAA 1272
```

RESULT 6
US-09-813-133A-3
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 55827
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-3

Query Match 30.4%; Score 478.6; DB 4; Length 55827;
Best Local Similarity 98.0%; Pred. No. 3.4e-106;
Matches 495; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

```
Qy       1051 TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGT 1110
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53334 TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGT 53393

Qy       1111 CATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTTAATTTACTGA 1170
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53394 CATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTTAATTTACTGA 53453

Qy       1171 TTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCCGTAGTTTTGATA 1230
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53454 TTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCCGTAGTTTTGATA 53513

Qy       1231 AAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTACTTTGCCATTAA 1290
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53514 AAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTACTTTGCCATTAA 53573

Qy       1291 GGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAAAAGTCTAGTTACC 1350
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      53574 GGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAAAATTGTAAAAGTCTAGTTACC 53632

Qy       1351 TACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTTTCGACGTTTGAC 1410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53633 TACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACTTTCGACGTTTGAC 53692

Qy       1411 TAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCATTGGATCCTACTCA 1470
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db      53693 TAGCCATCTCAAGCAAGTTTAATCAATGATCATCTCACGCTGATCATTGGATCCTACTCA 53752

Qy       1471 ACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAATTTTCAATAAATT 1530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53753 ACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAATTTTCAATAAATT 53812

Qy       1531 TCTTCTTCTCCTTTAAAAAAAAAAA 1555
              ||| ||| | | || | ||| | ||
Db      53813 TCTGCTTGTGCCTTTAGAAATACAA 53837
```

RESULT 7
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
; GENERAL INFORMATION:
; APPLICANT: Hartman, Jacob
; APPLICANT: Fulga, Netta
; APPLICANT: Mendelovitch, Simona
; APPLICANT: Gorecki, Marian
; TITLE OF INVENTION: PRODUCTION OF ENZYMATICALLY ACTIVE
; TITLE OF INVENTION: CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/782,760
; FILING DATE: 13-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/378,233
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0336/43847
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 927 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..927
US-08-782-760-5

Query Match 7.9%; Score 124; DB 2; Length 927;
Best Local Similarity 57.9%; Pred. No. 6.1e-21;
Matches 278; Conservative 0; Mismatches 195; Indels 7; Gaps 3;

```
Qy        613 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 672
              |||||||||||||| | ||| |||      ||    | |   ||| | ||   ||||| |
```

```
Db        362 GAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGACCCCA 421

Qy        673 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 732

Db        422 ACAGGAATTTTAATGC---TGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478

Qy        733 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 792

Db        479 AAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGATTTCA 538

Qy        793 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 852

Db        539 TCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAGATGA 598

Qy        853 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 912

Db        599 TGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAATGCCC 658

Qy        913 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 972

Db        659 TGGTGAAAGGTGCGGCAAAGGAGCTTGCC---ACTCTGCATGGCACCAAGTACACATATG 715

Qy        973 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1032

Db        716 GCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTATGATC 775

Qy       1033 TGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTC 1092

Db        776 AGGGAATCAAATATTCCTTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834


RESULT 8
PCT-US96-00995-5
; Sequence 5, Application PC/TUS9600995
;  GENERAL INFORMATION:
;    APPLICANT:  Bio-Technology General Corp.
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00995
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847-A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
PCT-US96-00995-5

  Query Match             7.9%;  Score 124;  DB 5;  Length 927;
  Best Local Similarity   57.9%;  Pred. No. 6.1e-21;
  Matches  278;  Conservative    0;  Mismatches  195;  Indels    7;  Gaps    3;

Qy        613 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 672

Db        362 GAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGACCCCA 421

Qy        673 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 732

Db        422 ACAGGAATTTTAATGC---TGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478

Qy        733 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 792

Db        479 AAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGATTTCA 538

Qy        793 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 852

Db        539 TCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAGATGA 598

Qy        853 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 912

Db        599 TGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAATGCCC 658

Qy        913 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 972

Db        659 TGGTGAAAGGTGCGGCAAAGGAGCTTGCC---ACTCTGCATGGCACCAAGTACACATATG 715

Qy        973 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1032

Db        716 GCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTATGATC 775

Qy       1033 TGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTC 1092

Db        776 AGGGAATCAAATATTCCTTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834


RESULT 9
US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..921
US-08-696-139-3

  Query Match             7.8%;  Score 123.4;  DB 1;  Length 921;
  Best Local Similarity   56.3%;  Pred. No. 8.4e-21;
  Matches  273;  Conservative    0;  Mismatches  206;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |     | | || ||||||||||||||| |||
Db        321 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCG 380

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
               |||  | ||||    | |   ||||| || |||||||  || || || ||||   |
Db        381 CTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTTGATGCTGG 440

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
              |||||       || || || |    | | |||   || || ||||||||||  |
Db        441 G---TGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGC 497

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
              | | ||||| |||  |||    |||||  |||||  |||  |  | |  ||  ||  |
Db        498 TGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTC 557

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || |||||||||  | |  || || |||||||| |||   ||| | |  || ||||||||
Db        558 CATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTA 617

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              |      |       | |  || |||  ||  ||  |    | || |  |  || ||
Db        618 TGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAA 677

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |   |||      ||      |   |||| ||| ||| | ||||  ||  |   |||  |
Db        678 AGAACTTG---CTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAAT 734

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
              || |  ||| |||| ||   || || |||   |||||    || ||||||||||| ||
Db        735 CTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTT 794

Qy       1052 TACAT 1056
              || |
Db        795 CACCT 799


RESULT 10
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1215 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1215
US-08-696-139-1

  Query Match             7.8%;  Score 123.4;  DB 1;  Length 1215;
  Best Local Similarity   56.3%;  Pred. No. 9.2e-21;
  Matches  273;  Conservative    0;  Mismatches  206;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |     | | || ||||||||||||||| |||
Db        615 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCG 674

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
               |||  | ||||    | |   ||||| || |||||||  || || || ||||   |
Db        675 CTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTTGATGCTGG 734

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
              |||||       || || || |    | | |||   || || ||||||||||  |
Db        735 G---TGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGC 791

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
              | | ||||| |||  |||    |||||  |||||  |||  |  | |  ||  ||  |
Db        792 TGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTC 851

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || |||||||||  | |  || || |||||||| |||   ||| | |  || ||||||||
Db        852 CATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTA 911

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              |      |       | |  || |||  ||  ||  |    | || |  |  || ||
Db        912 TGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAA 971

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |   |||      ||      |   |||| ||| ||| | ||||  ||  |   |||  |
Db        972 AGAACTTG---CTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAAT 1028

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
              || |  ||| |||| ||   || || |||   |||||    || ||||||||||| ||
```

```
Db       1029 CTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTT 1088

Qy       1052 TACAT 1056
                 || |
Db       1089 CACCT 1093


RESULT 11
US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 38:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1263 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

  Query Match             6.9%;  Score 109;  DB 4;  Length 1263;
  Best Local Similarity   54.4%;  Pred. No. 2.8e-17;
  Matches  264;  Conservative    0;  Mismatches  215;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |      | | |  ||| | ||||||||||  ||
Db        645 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 704

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
                ||   | || |      |   ||||| || ||||||| || || || ||||   |
Db        705 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 761

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
                  ||||||||    || || ||  |   | | ||    ||||| ||||||||||| |
Db        762 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 821

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
              | |||||| |||   |||   |||||  |||||  |||| |  | |  ||  ||
Db        822 CGCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 881

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || || |||||  | |  || || || |||||||||    ||  | |  || || || ||
Db        882 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 941

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              | |    |      |  |  || ||| ||  ||  |    | || |  |   | ||
Db        942 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 1001

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |  ||| |    |     |    |||| ||| ||| ||||||  ||  |   |||  |
Db       1002 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 1058

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
              || |  ||| |||| ||    ||||| |||   |||||    || |||| |||||| ||
Db       1059 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 1118

Qy       1052 TACAT 1056
                 || |
Db       1119 CACCT 1123


RESULT 12
US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  999 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-67
```

```
  Query Match             6.8%;  Score 107.4;  DB 2;  Length 999;
  Best Local Similarity   54.2%;  Pred. No. 6.4e-17;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
              ||   | ||  |       |   ||||| || |||||||  || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
              ||||||||    || || ||  |   | | ||    ||||| |||||||||| |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
              | |||||  |||   ||    |||||  |||||  ||||  |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              | | .  |    |  |  || |||  ||  ||  |    | || |  |   | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |   ||| |    |     |  ||||| |||  ||| |||||  ||  |   |||  |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
              || |  ||| |||| ||   |||||  |||   |||||    || |||| |||||| ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1052 TACAT 1056
              ||| |
Db        861 CACCT 865


RESULT 13
US-09-011-769A-50
; Sequence 50, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 999 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION: 1..987
;        FEATURE:
;             NAME/KEY:  mat_peptide
;             LOCATION: 67..987
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-011-769A-50

  Query Match             6.8%;  Score 107.4;  DB 4;  Length 999;
  Best Local Similarity   54.2%;  Pred. No. 6.4e-17;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
              ||   | ||  |       |   ||||| || |||||||  || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
              ||||||||    || || ||  |   | | ||    ||||| |||||||||| |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
              | |||||  |||   ||    |||||  |||||  ||||  |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              | | .  |    |  |  || |||  ||  ||  |    | || |  |   | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |   ||| |    |     |  ||||| |||  ||| |||||  ||  |   |||  |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
              || |  ||| |||| ||   |||||  |||   |||||    || |||| |||||| ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1052 TACAT 1056
              ||| |
Db        861 CACCT 865


RESULT 14
US-08-860-882A-64
; Sequence 64, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
```

```
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1053 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-64

  Query Match             6.8%;  Score 107.4;  DB 2;  Length 1053;
  Best Local Similarity   54.2%;  Pred. No. 6.5e-17;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
              ||   | || |       |   ||||| || |||||||  || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
              ||||||||    || || ||  |   | | ||    ||||| |||||||||| |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
                | ||||| |||   ||    |||||  |||||  |||| |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              | |   |    |  |  || |||  ||  ||  |   | || | |   | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |   ||| |    |     |    |||| ||| ||| ||||||  ||  |   |||  |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
```

```
              || |  ||| |||| ||    ||||| |||   |||||    || |||| |||||| ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1052 TACAT 1056
              || |
Db        861 CACCT 865


RESULT 15
US-09-011-769A-46
; Sequence 46, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1053 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION: 1..1047
;       FEATURE:
;            NAME/KEY:  mat_peptide
;            LOCATION: 67..1047
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-011-769A-46

  Query Match             6.8%;  Score 107.4;  DB 4;  Length 1053;
  Best Local Similarity   54.2%;  Pred. No. 6.5e-17;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        572 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 631
              | ||  |  || |  || | | |     |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        632 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 691
              ||   | || |       |   ||||| || |||||||  || || || ||||   |
```

```
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        692 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 751
              ||||||||     || || ||  |    | | ||     ||||| |||||||||| |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        752 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 811
              | |||||  |||    ||     |||||  |||||  |||| |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        812 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 871
              || || |||||  | |  || || || ||||||||     ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        872 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 931
              | |    |       |  |  || |||  ||  ||  |    | || |  |   | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        932 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 991
              |   ||| |    |     |    |||| ||| ||| ||||||  ||  |   |||  |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        992 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1051
              || |  ||| |||| ||    ||||| |||   |||||    || |||| |||||| ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1052 TACAT 1056
              || |
Db        861 CACCT 865
```

Search completed: January 1, 2004, 15:40:55
Job time : 118.353 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 13:56:00 ; Search time 541.923 Seconds
(without alignments)
10046.789 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1372	87.2	1728	10	US-09-880-107-2396	Sequence 2396, Ap
	2	858.8	54.6	1272	13	US-10-379-836-1	Sequence 1, Appli
	3	853.8	54.3	1400	9	US-09-925-302-24	Sequence 24, Appl
	4	331.2	21.1	416	10	US-09-960-352-14595	Sequence 14595, A
	5	286.8	18.2	397	10	US-09-960-352-8425	Sequence 8425, Ap
c	6	137.2	8.7	431	10	US-09-917-800A-468	Sequence 468, App
	7	131	8.3	1547	13	US-10-027-632-265133	Sequence 265133,
	8	131	8.3	1547	13	US-10-027-632-265134	Sequence 265134,
	9	131	8.3	1547	13	US-10-027-632-265135	Sequence 265135,
	10	131	8.3	1547	13	US-10-027-632-265136	Sequence 265136,
	11	131	8.3	1547	14	US-10-027-632-265133	Sequence 265133,
	12	131	8.3	1547	14	US-10-027-632-265134	Sequence 265134,
	13	131	8.3	1547	14	US-10-027-632-265135	Sequence 265135,
	14	131	8.3	1547	14	US-10-027-632-265136	Sequence 265136,
	15	107.6	6.8	1254	15	US-10-229-546-3	Sequence 3, Appli
	16	107.6	6.8	1622	12	US-10-429-802-21	Sequence 21, Appl
	17	107.6	6.8	1622	15	US-10-229-546-1	Sequence 1, Appli
	18	107.6	6.8	1633	13	US-10-341-434-187	Sequence 187, App
	19	101	6.4	1870	10	US-09-910-059-112	Sequence 112, App
	20	101	6.4	2154	10	US-09-910-059-124	Sequence 124, App
	21	94.6	6.0	1332	10	US-09-954-456-1141	Sequence 1141, Ap
	22	92.8	5.9	1311	14	US-10-200-344-9	Sequence 9, Appli
	23	92.8	5.9	1993	12	US-10-274-639-33	Sequence 33, Appl
	24	92.8	5.9	2128	14	US-10-200-344-13	Sequence 13, Appl
	25	91.6	5.8	936	15	US-10-106-698-1866	Sequence 1866, Ap
	26	88.6	5.6	517	15	US-10-198-846-12472	Sequence 12472, A
	27	88.6	5.6	1050	14	US-10-200-344-11	Sequence 11, Appl
	28	87.6	5.6	1200	15	US-10-200-910-7	Sequence 7, Appli
	29	83.6	5.3	945	15	US-10-200-910-3	Sequence 3, Appli
	30	83.6	5.3	945	15	US-10-200-910-11	Sequence 11, Appl
	31	80.4	5.1	451	15	US-10-060-036-3265	Sequence 3265, Ap
	32	80	5.1	1125	9	US-09-888-615-2	Sequence 2, Appli
	33	80	5.1	1332	16	US-10-176-306-75	Sequence 75, Appl
	34	80	5.1	1603	16	US-10-176-306-73	Sequence 73, Appl
	35	80	5.1	1826	13	US-10-252-157-453	Sequence 453, App
	36	79.2	5.0	878	15	US-10-198-846-4746	Sequence 4746, Ap
	37	78.4	5.0	741	14	US-10-200-344-5	Sequence 5, Appli
c	38	78.4	5.0	1603	16	US-10-176-306-73	Sequence 73, Appl
	39	76.8	4.9	948	9	US-09-888-615-1	Sequence 1, Appli
	40	76.8	4.9	1187	15	US-10-198-846-13457	Sequence 13457, A
	41	72.8	4.6	1341	9	US-09-925-297-69	Sequence 69, Appl
	42	71.6	4.6	1306	9	US-09-923-779-144	Sequence 144, App
	43	71.6	4.6	1306	9	US-09-962-436-557	Sequence 557, App
	44	70.6	4.5	1254	9	US-09-925-297-58	Sequence 58, Appl
	45	70.6	4.5	1260	13	US-10-345-680-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Query Match 87.2%; Score 1372; DB 10; Length 1728;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 1558; Conservative 0; Mismatches 5; Indels 163; Gaps 2;

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 122
```

```
Qy       121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
Db       123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy       181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
Db       183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 242

Qy       241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
Db       243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy       301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
Db       303 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy       361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
Db       363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy       421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
Db       423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy       481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
Db       483 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 542

Qy       541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
Db       543 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 602

Qy       601 TAGGCCAT---------------------------------------------------- 608
Db       603 TAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTG 662

Qy       609 ----------------------------------------------------------A 609
Db       663 TGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGA 722

Qy       610 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 669
Db       723 ATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACC 782

Qy       670 TGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 729
Db       783 TGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCT 842

Qy       730 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 789
Db       843 CGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTT 902

Qy       790 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 849
Db       903 TCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGC 962

Qy       850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
Db       963 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 1022

Qy       910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
Db      1023 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACAC 1082

Qy       970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
Db      1083 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1142

Qy      1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
Db      1143 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 1202

Qy      1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
Db      1203 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

Qy      1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
Db      1263 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1322

Qy      1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
Db      1323 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1382

Qy      1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
Db      1383 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1442

Qy      1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
Db      1443 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1502

Qy      1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
Db      1503 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1562

Qy      1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
Db      1563 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1622

Qy      1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
Db      1623 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1682

Qy      1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1563
Db      1683 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAA 1728
```

RESULT 2
US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 1272
; TYPE: DNA
; ORGANISM: Papio hamadryas
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1269)
US-10-379-836-1

Query Match 54.6%; Score 858.8; DB 13; Length 1272;
Best Local Similarity 83.9%; Pred. No. 5.8e-208;
Matches 1067; Conservative 0; Mismatches 42; Indels 163; Gaps 2;

```
Qy        18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77
Db         1 ATGAAGCTTTGCAGTCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy        78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137
Db        61 TTCGCGTTTCAGAGTGGCCAGGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
```

```
Qy       138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197
Db       121 GTGCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCGGAC 180

Qy       198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257
Db       181 CTTATTGAGAAGAAAAAACAAGTCCATTTTTTTGTAAATTCATCTGATGTCGACAATGTG 240

Qy       258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317
Db       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCCTGCTGGCAGATGTGGAAGAT 300

Qy       318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377
Db       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy       378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437
Db       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAACTTATAACTGAGAAGTAT 420

Qy       438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497
Db       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCCTATGAGAAGCACCCACTTTATGTT 480

Qy       498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557
Db       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATGTGGATTGACTGTGGAATC 540

Qy       558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 608
Db       541 CATGCCAGAGAATGGATCTCCCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTGAA 600

Qy       609 ------------------------------------------------------------ 608

Db       601 TACTACGGGATAATAGGGGAATATACCAATCTTCTGAGGCATGTGGATTTCTATGTTATG 660

Qy       609 ----------------------------------------AATCGAATGTGGAGAAAG 626
                                                    ||||||||||||||||||
Db       661 CCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy       627 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 686
Db       721 AACCGTTCTTTCTATGCGAACAATCGTTGCATCGGAACAGACCTGAACAGGAACTTTGCG 780

Qy       687 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 746
Db       781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTTCTCATGCTCGGAAACCTACTGTGGA 840

Qy       747 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 806
Db       841 CTTTATCCTGAGTCAGAACCAGAAGCGAAGGCGGTGGCTAATTTCTTGAGAAGAAATATC 900

Qy       807 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 866
Db       901 AACCACATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATCGTGTTTCCATAT 960

Qy       867 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 926
Db       961 TCCTATACTCGAAGCAAAAGCAAAGACCACGAGGAATTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy       927 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 986
Db      1021 GTTCGTGCTATTCAGAAAACCAGTAAAAATATCAGGTATACACATGGCCGTGGCTCAGAA 1080

Qy       987 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1046
Db      1081 ACCTTATACCTAGCTCCTGGAGGTGCGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy      1047 TCGT----------------------------------------------------TTAC 1054
              ||||                                                    ||||
Db      1141 TCGTTTACAATTGAACTTCGAGATACGGGCAAATACGGATTCTTGCTGCCTGAGCGTTAC 1200

Qy      1055 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1114
              |||||||||||| |||| ||| ||||||||||||||||||||||||||||||||||||||
Db      1201 ATCAAACCCACTTGTAAAGACGCTTTTGCCGCTGTCTCTAAAATAGCTTGGCATGTCATT 1260

Qy      1115 AGGAATGTTTAA 1126
              ||||||||||||
Db      1261 AGGAATGTTTAA 1272
```

RESULT 3
US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Query Match 54.3%; Score 853.8; DB 9; Length 1400;
Best Local Similarity 93.3%; Pred. No. 1.2e-206;
Matches 934; Conservative 1; Mismatches 13; Indels 53; Gaps 2;

```
Qy       607 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 666
             | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       177 AGAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 236

Qy       667 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 726
             |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       237 ACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 296

Qy       727 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 786
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       297 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 356

Qy       787 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 846
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       357 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 416

Qy       847 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 906
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       417 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 476

Qy       907 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATA 966
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db       477 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATA 536

Qy       967 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 1026
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       537 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 596

Qy      1027 ATGATTTGGGCATCAAATATTCGT---------------------------------- 1050
             |||||||||||||||||||||||
Db       597 ATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGAT 656

Qy      1051 ----------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTA 1094
                             ||||||||||||||||||||||||||||||||||||||||||||
Db       657 TCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTA 716
```

```
Qy       1095 AAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCG 1154
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        717 AAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCG 776

Qy       1155 TATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTT 1214
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        777 TATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTT 836

Qy       1215 ATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGT 1274
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        837 ATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGT 896

Qy       1275 GCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTG 1334
             |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        897 GCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAAAATTG 955

Qy       1335 TAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCA 1394
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        956 TAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCA 1015

Qy       1395 ACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGAT 1454
             ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db       1016 ACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAWGATCATCTCACGCTGAT 1075

Qy       1455 CATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCC 1514
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1076 CATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCC 1135

Qy       1515 AAATTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAA 1555
             |||||||||||||||||| ||| | | || | ||| | ||
Db       1136 AAATTTTCAATAAATTTCTGCTTGTGCCTTTAGAAATACAA 1176
```

RESULT 4
US-09-960-352-14595
; Sequence 14595, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14595
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
US-09-960-352-14595

Query Match 21.1%; Score 331.2; DB 10; Length 416;
Best Local Similarity 87.3%; Pred. No. 6e-74;
Matches 363; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        167 GATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTT 226
              |||||||||||||||||| ||| |||||||   |||||||||  | |  ||||||||||
Db          1 GATTGTTCTCTGGCAGCCAGTAGCAGCTGAATATATTGTGAAGGGATACGAAGTCCATTT 60

Qy        227 TTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCC 286
              ||||||||||||||||||   ||||||||||||||||||||||| |||| |||||||
Db         61 TTTTGTGAATGCATCTGATGTAAGCAATGTGAAAGCCCATTTAAATGCGAGCAGAATTCC 120

Qy        287 ATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACAC 346
              || ||| ||| || ||| | | ||||||||||||||| |  |||||||  |||||| |||||
Db        121 ATTCAGGGTCCTGGTGGAAAATGTGGAAGATCTTATCCGGCAGCAGACTTCCAATGACAC 180

Qy        347 AGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTA 406
```

```
Db        181 CATCAGCCCCCGGGCATCCTCCTCCTACTATGAACAGTATCACTCACTAAATGAGATCTA 240

Qy        407 TTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGG 466
              |||||||||||| |||| |||||| || |||||||||| ||  |||||||||||||||
Db        241 TTCTTGGATAGAAGTTATGACTGAGCGGTATCCTGATATGGTTGAAAAAATCCACATTGG 300

Qy        467 ATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGC 526
              ||||||||  |||||||||||||| |||||||||||||||||||  ||||||||||  ||
Db        301 ATCCTCATACGAGAAGTACCCACTTTATGTTTTAAAGGTTTCTAAAAAAGAACAAAGGGC 360

Qy        527 CAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTG 582
              |||||||||||| |||||||||||||||||||| |||||||| ||||||||||||
Db        361 CAAAAATGCCATGTGGATTGACTGTGGAATCCACGCCAGAGAGTGGATCTCTCCTG 416
```

RESULT 5
US-09-960-352-8425
; Sequence 8425, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8425
; LENGTH: 397
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 36-LIB34-005-Q1-E1-A12
US-09-960-352-8425

Query Match 18.2%; Score 286.8; DB 10; Length 397;
Best Local Similarity 83.9%; Pred. No. 1.2e-62;
Matches 324; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

```
Qy         12 GTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAG 71
              || |||||||||||||  ||| | | |||||||||  |||  || || |||||||| | ||
Db         12 GTGGGGATGAAGCTTTATAGCTTGGGAGTCCTTGTCGCCACCGTCCTGTTCTGTGGGGAG 71

Qy         72 CATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAA 131
              |||| |||||| || ||||| ||||| ||| || ||||||||||||||||||| ||||||
Db         72 CATGCCTTCGCCTTTCAGAGGGGCCAGGTTTTATCTGCTCTTCCTAGAACCTCCAGGCAA 131

Qy        132 GTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACA 191
              |||||| |||| |||||| |||||||||||||  |||||||||||||||||||| ||| ||
Db        132 GTTCAAATTCTGCAGAATGTTACTACAACATACAAGATTGTTCTCTGGCAGCCAGTAGCA 191

Qy        192 GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGAC 251
              |||||   ||||||||||  | |  |||||||||||||||| ||||||||||||||   |
Db        192 GCTGAATATATTGTGAAGGGATACGAAGTCCATTTTTTTGTGAATGCATCTGATGTAAGC 251

Qy        252 AATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTG 311
              |||||||||||||||||||||| |||| |||||||| ||| ||| || ||| | | |||
Db        252 AATGTGAAAGCCCATTTAAATGCGAGCAGAATTCCATTCAGGGTCCTGGTGGAAAATGTG 311

Qy        312 GAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCG 371
              |||||||||| |  ||||||| |||||| |||||  |||||||||| || ||| | ||
Db        312 GAAGATCTTATCCGGCAGCAGACTTCCAATGACACCATCAGCCCCCGGGCATCCTCCTCC 371

Qy        372 TACTATGAACAGTATCACTCACTAAA 397
              |||||||||||||||||||||||||
Db        372 TACTATGAACAGTATCACTCACTAAA 397
```

RESULT 6
US-09-917-800A-468/c

```
; Sequence 468, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 468
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA998857
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(431)
;   OTHER INFORMATION: n = a or c or g or t
US-09-917-800A-468

  Query Match             8.7%;  Score 137.2;  DB 10;  Length 431;
  Best Local Similarity   67.5%;  Pred. No. 1.4e-24;
  Matches  268;  Conservative    0;  Mismatches   69;  Indels   60;  Gaps    3;

Qy        894 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 953
              ||||||||||||||| |||| ||||| |||||||||||||| ||||| |   | || ||||
Db        431 CATGAGGAACTGTCTNTAGTGGCCAGCGAAGCAGTTCGTGCCATTGAAAGTATTAATAAA 372

Qy        954 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 1013
              || |||||||| |||||||||  |||||||||||  ||||| ||||||||||||||||
Db        371 AACACCAGGTACACACATGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCT 312

Qy       1014 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTA---------------------- 1053
              || ||||||||||||||||||||||||||||||||||||
Db        311 GATGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACGATTGAACTTCGGGATACA 252

Qy       1054 -------------------------------CATCAAACCCACCTGTAGAGAAGCTTTT 1081
                                               |||||||||||| ||   |||||||||
Db        251 GGCAGATACGGGTTCTTGCTGCCTGAGAGATTCATCAAACCCACTTGCGCAGAAGCTTTG 192

Qy       1082 GCCGCTGTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTAT 1141
              |||||  |||||||||||||||||||||||||| |||||    ||||  |||       ||
Db        191 GCCGCAGTCTCTAAAATAGCTTGGCATGTCATCAGGAACAGTTAACACCC------TTTC 138

Qy       1142 CATTCTGCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT- 1200
              |  |   ||      ||||| |||| ||||| |||||| || ||||  |||  |  || |
Db        137 CTCTGCTCTCATTACTTTTATTTTATTGATTTCAGCAACACTAAATTGTTGCACTAGCTT 78

Qy       1201 -TATTTTTAAGTTTTATCCGTAGTTTTGATAAAAGAT 1236
```

```
              ||  |||||    | ||| | |||||| | ||  ||
Db         77 CTAAGTTTAATCAGTTTCCTTGGTTTTGTTGAAGAAT 41


RESULT 7
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265133
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265133

  Query Match             8.3%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 8
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265134
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134

  Query Match             8.3%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 9
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265135
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265135

  Query Match             8.3%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 10
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265136
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

  Query Match             8.3%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547
```

```
RESULT 11
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265133
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265133

  Query Match             8.3%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 12
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265134
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134

  Query Match             8.3%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 13
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265135
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265135

  Query Match             8.3%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 14
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265136
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

  Query Match             8.3%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        702 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        762 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        822 TACATCAGCAT 832
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 15
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1254
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1254)
US-10-229-546-3

  Query Match             6.8%;  Score 107.6;  DB 15;  Length 1254;
  Best Local Similarity   55.6%;  Pred. No. 8.9e-17;
  Matches  250;  Conservative    0;  Mismatches  194;  Indels    6;  Gaps    2;

Qy        607 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 666
              | || || |||||||||||| || |||||     |     |||     | ||||||| || |
Db        686 AGAACCGCATGTGGAGAAAAAATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTG 745

Qy        667 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 726
              |||| || || || |||    |   |   |||           |    ||| |  | |||
Db        746 ACCTCAACAGGAATTTTAATGCTTCA---TGGAACTCCATTCCTAACACCAATGACCCAT 802

Qy        727 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 786
              |  | || | |||  | ||   |   || ||||| ||   |||   ||| || ||  |||
Db        803 GTGCAGATAACTATCGGGGCTCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTA 862

Qy        787 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 846
              |||| | |||||  | | ||  | || || | |||||||| | | ||||| |||||||
Db        863 ATTTCATTAGAAGCCACCTGAATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCC 922

Qy        847 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 906
              ||    || ||||||| |||   ||||||  ||           |||||||||  ||
Db        923 AGATGCTATTGTTTCCCTATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGG 982

Qy        907 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATA 966
              |   ||| || |   |    |   || | ||     |||| |  |  | ||| | || |
Db        983 CCAAAGTTGCAA---AGATTGGCACTGATGTTCTATCAACTCGATATGAAACCCGCTACA 1039

Qy        967 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 1026
              ||||||        | ||| | ||||       | ||   |       || |||   |
Db       1040 TCTATGGCCCAATAGAATCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTT 1099

Qy       1027 ATGATTTGGGCATCAAATATTCGTTTACAT 1056
              ||||  |||||||||||  |  | ||| | |
Db       1100 ATGACCTGGGCATCAAACACACATTTGCCT 1129


Search completed: January  1, 2004, 18:40:36
Job time : 561.923 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 10:44:24 ; Search time 3636.93 Seconds
(without alignments)
10511.864 Million cell updates/sec

Title: US-09-980-881A-1
Perfect score: 1573
Sequence: 1 agaaaattgctgttgggatg.....aaaaaaaaaaaaaaaaaaa 1573

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	717.8	45.6	973	9	AL536068	AL536068 AL536068
	2	690.8	43.9	932	13	BX451532	BX451532 BX451532
	3	663.2	42.2	971	9	AL536069	AL536069 AL536069
	4	656	41.7	1446	11	AK004045	AK004045 Mus muscu
	5	622.8	39.6	780	10	BG616950	BG616950 602615625
	6	615.6	39.1	806	10	BG567264	BG567264 602589745
	7	604.2	38.4	788	10	BG618629	BG618629 602646064
	8	594.8	37.8	702	2	HSM089979	Bx496754 Homo sapi
	9	581.2	36.9	747	10	BG566051	BG566051 602582652
	10	579.2	36.8	725	10	BG618239	BG618239 602645967
	11	561.8	35.7	738	10	BG616456	BG616456 602614618
	12	555.2	35.3	735	9	AV646979	AV646979 AV646979
	13	536.2	34.1	541	14	CB157619	CB157619 K-EST0216
	14	532.2	33.8	750	10	BG618813	BG618813 602646186
	15	523.8	33.3	691	10	BG568240	BG568240 602587149
	16	510.4	32.4	623	9	AV694058	AV694058 AV694058
	17	509.4	32.4	583	9	AV693037	AV693037 AV693037
	18	509.4	32.4	602	9	AV692032	AV692032 AV692032
	19	507.4	32.3	621	9	AV698425	AV698425 AV698425
	20	501.2	31.9	601	9	AV658390	AV658390 AV658390
	21	491.6	31.3	643	9	AV688323	AV688323 AV688323
	22	483.4	30.7	485	13	BX091627	BX091627 BX091627
c	23	468	29.8	589	9	AV719484	AV719484 AV719484
	24	466	29.6	558	9	AV662298	AV662298 AV662298
c	25	465.4	29.6	611	14	N63924	N63924 za28c09.s1
	26	465	29.6	603	10	AW950700	AW950700 EST362770
c	27	464	29.5	586	9	AV719308	AV719308 AV719308
	28	461.8	29.4	688	9	AV651709	AV651709 AV651709
c	29	460.2	29.3	495	9	AI271641	AI271641 qi19h11.x
c	30	460.2	29.3	585	9	AV720894	AV720894 AV720894
c	31	460.2	29.3	611	9	AA700377	AA700377 zj73b12.s
	32	457.8	29.1	537	14	N98450	N98450 za28c09.r1
	33	455	28.9	577	14	CB154355	CB154355 K-EST0212
	34	450.4	28.6	462	9	AV681868	AV681868 AV681868
	35	444.2	28.2	563	14	CB162759	CB162759 K-EST0223
c	36	442	28.1	576	9	AV720000	AV720000 AV720000
	37	441.6	28.1	623	9	AV658178	AV658178 AV658178
c	38	438.6	27.9	586	9	AV719997	AV719997 AV719997
	39	436	27.7	510	14	CB156851	CB156851 K-EST0215
	40	432.8	27.5	1004	14	BY704293	BY704293 BY704293
	41	427.4	27.2	455	14	W88434	W88434 zh69f04.r1
c	42	422.4	26.9	593	14	W89198	W89198 zh69f04.s1
	43	420.6	26.7	747	10	BG565100	BG565100 602583720
c	44	418.4	26.6	450	9	AI076434	AI076434 oz17b02.x
	45	417.4	26.5	958	9	AI322676	AI322676 mj24d06.y

ALIGNMENTS

RESULT 1
AL536068/c
LOCUS AL536068 973 bp mRNA linear EST 31-MAY-2003
DEFINITION AL536068 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF022YE21 3-PRIME, mRNA sequence.
ACCESSION AL536068
VERSION AL536068.2 GI:31260938
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 973)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799561.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6944.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF022AC11NP1&cluster=6944.r. Contact : Feng Liang Email : fliang@lifetech.com URL :

```
          http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0DF022AC11NP1.
FEATURES             Location/Qualifiers
     source          1. .973
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE21"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      284 a    193 c    188 g    250 t     58 others
ORIGIN

  Query Match            45.6%;  Score 717.8;  DB 9;  Length 973;
  Best Local Similarity  86.0%;  Pred. No. 2.5e-77;
  Matches  793;  Conservative   44;  Mismatches   31;  Indels   54;  Gaps    3;

Qy        607 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 666
              | |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        956 AGAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATTGGAACAG 897

Qy        667 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 726
              |||||||||  |||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        896 ACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 837

Qy        727 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 777

Qy        787 GTTTCTTGAGAAGAAATATCAA-CCAGATTAAAGCATACATCAGCATGCATTCATACTCC 845
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        776 GTTTCTTGAGAAGAAATATCAAGCCAGATTAAAGCATACATCAGCATGCATTCATACTCC 717

Qy        846 CAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTG 905
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        716 CAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTG 657

Qy        906 TCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTAT 965
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        656 TCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTAT 597

Qy        966 ACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATC 1025
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        596 ACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATC 537

Qy       1026 TATGATTTGGGCATCAAATATTCGTTTACA---------------------------- 1055
              |||||||||||||||| | ||||||||||
Db        536 TATGATTTGGGCATCAATTTTTCGTTTACAATTGAACTTCGAGTTACGGGCACATACGGA 477

Qy       1056 ----------------------TCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCT 1093
                                   |||| ||| ||||||||||||||||||||||||||||||
Db        476 TTCTTGCTGCCGGGGCGTTTCCTCAACCCCCCCTGTAGAGAAGCTTTTGCCGCTGTCTCT 417

Qy       1094 AAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCC 1153
              |||||:|||||||||| |:|||||||||: :||::|:|||| |||||: ||||| ||  ||
Db        416 AAAATMGCTTGGCATGCCMTTAGGAATGYCYAAYSCMCCTGGTTTTVCCATTCSGCCCCC 357

Qy       1154 GTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTT 1213
              |||||||||||||:|||:||:|||||:::|| |||||||:||:|||||||||| |||||
Db        356 GTATTTTAATTTACKGATKCCVGCAAGRSSAATTCATTGTWTCRGATTATTTTTGAGTTT 297

Qy       1214 TATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAG 1273
              |:|||||:||||||:|:||||:||||: | ||||||||:||||:|||||| |||||||
Db        296 TWTCCGTWGTTTTGWTWAAAGWTTTTYTTTTTCCTTGGTTYTGTCWGAGAACTTAATAAG 237

Qy       1274 TGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATT 1333
              ||||::|||||||||||||:||||:||||||||||:|||||||||||| :|||||||||
```

```
Db        236 TGCTWYTTTGCCATTAAGGCWGACTWGGGTTCATGTYTTTTTACCCTTT-WAAAAAAATT 178

Qy       1334 GTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGC 1393
              :|:|||||||:|||:|||||||||||||||||||||||||||||||||||||:|:::|
Db        177 KTWAAAGTCTWGTTWCCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCYCMMSC 118

Qy       1394 AACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGA 1453
                 :::|:| |||||||||||||||:|||||||||||||||| ||||:||||||||||||
Db        117 SGSKKTSGGCGTTTGACTAGCCATYTCAAGCAAGTTTAATCAATGATCMTCTCACGCTGA 58

Qy       1454 TCATTGGATCCTACTCAACAAA 1475
              |||||||||||||  |  |  ||
Db         57 TCATTGGATCCTANTNNAACAA 36


RESULT 2
BX451532
LOCUS       BX451532                932 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX451532 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION   BX451532
VERSION     BX451532.1  GI:31024255
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6944.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0BAF016ZB11_AF01523_1&cluster=6944.r.
            Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0BAF016ZB11_AF01523_1.
FEATURES             Location/Qualifiers
     source          1. .932
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE21"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      281 a    189 c    170 g    283 t      9 others
ORIGIN

  Query Match            43.9%;  Score 690.8;  DB 13;  Length 932;
  Best Local Similarity  91.6%;  Pred. No. 4.5e-74;
  Matches  785;  Conservative    0;  Mismatches   18;  Indels   54;  Gaps    3;

Qy        720 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 779
              | |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         14 TTCCTATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 73

Qy        780 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 839
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         74 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 133

Qy        840 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 899
```

```
Db        134 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 193

Qy        900 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959

Db        194 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACC 253

Qy        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019

Db        254 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 313

Qy       1020 TGGATCTATGATTTGGGCATCAAATATTCGT---------------------------- 1050

Db        314 TGGATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACA 373

Qy       1051 ----------------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 1087

Db        374 TACGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 433

Qy       1088 GTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCT 1147

Db        434 GTCTCTAAAATAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCT 493

Qy       1148 GCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTT 1207

Db        494 GCTTCCGTATTTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTT 553

Qy       1208 AAGTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCT 1267

Db        554 AAGTTTTATCCGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCT 613

Qy       1268 AATAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAA 1327

Db        614 AATAAGTGCTACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTNTACCCTTTAAAAA 673

Qy       1328 AAAATTGTAAAAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATC 1387

Db        674 AAAA-TGTAAAAGCCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATC 732

Qy       1388 TCAAGCAACTTTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCA 1447

Db        733 TCAAGCAACTTTCGACGGTTGACTAGCCATCCTAAGCCAGGTTAATCAATGATCATCTCA 792

Qy       1448 CGCTGATCATTGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTT 1507

Db        793 CGCTGATCATTGGATCCTACTCAAC-AAAGGAAGGGTGGGCAGGAGTACATTAAAGATTT 851

Qy       1508 CTGCTCCAAATTTTCAA 1524

Db        852 CTGGTTCCAAATTTCAA 868


RESULT 3
AL536069
LOCUS       AL536069                971 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION   AL536069
VERSION     AL536069.2  GI:31260939
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12799562.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6944.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF022AC11QP1&cluster=6944.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF022AC11QP1.
FEATURES             Location/Qualifiers
     source          1...971
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE21"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      282 a    201 c    211 g    272 t      5 others
ORIGIN

  Query Match           42.2%;  Score 663.2;  DB 9;  Length 971;
  Best Local Similarity   86.8%;  Pred. No. 9e-71;
  Matches  786;  Conservative    3;  Mismatches    6;  Indels  111;  Gaps    1;

Qy         15 GGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCAT 74

Db         62 GGGAWGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCAT 121

Qy         75 GTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTT 134

Db        122 GTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTT 181

Qy        135 CAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCT 194

Db        182 CAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCT 241

Qy        195 GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT 254

Db        242 GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT 301

Qy        255 GTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAA 314

Db        302 GTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAA 361

Qy        315 GATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC 374

Db        362 GATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC 421

Qy        375 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 434

Db        422 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 481

Qy        435 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 494

Db        482 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 541

Qy        495 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 554

Db        542 GTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGA 601

Qy        555 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT------ 608

Db        602 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACT 661

Qy        609 ---------------------------------------------------------- 608

Db        662 CAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTT 721
```

```
Qy        609 ----------------------------------------------AATCGAATGTGGAGA 623
                                                           ||||||||||||||||
Db        722 ATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGA 781

Qy        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683
              |||||||||||||||||||||||||||||| |||||||||||||||| ||||||
Db        782 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATTGGAACAGACCTGAATAGGAACTTT 841

Qy        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743
              |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 GCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 901

Qy        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db        902 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAW 961

Qy        804 ATCAAC 809
              ||:|||
Db        962 ATMAAC 967
```

RESULT 4
AK004045
LOCUS AK004045 1446 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1110032P04 product:carboxypeptidase B2 (plasma), full insert sequence.
ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1446)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1446
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1110032P04"
/db_xref="MGI:1900711"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
CDS 42. .1310
/note="unnamed protein product; carboxypeptidase B2 (plasma) (MGD|MGI:1891837, GB|NM_019775, evidence: BLASTN, 99%, match=1419)
putative"
/codon_start=1
/protein_id="BAB23141.1"
/db_xref="GI:12835068"

```
                     /db_xref="MGI:1891837"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      421 a    329 c    311 g    385 t
ORIGIN

  Query Match           41.7%;  Score 656;  DB 11;  Length 1446;
  Best Local Similarity  72.2%;  Pred. No. 5.1e-70;
  Matches 1009;  Conservative    0;  Mismatches  215;  Indels  174;  Gaps    5;

Qy          4 AAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCT 63
              ||  |  |||||||||||||||||    |||||| | |||| ||| ||||  | ||||
Db         28 AAGTCACTGTTGGGATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTA-- 85

Qy         64 GTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCT 123
               |||||||||||| |||||| || |||||||||| ||| || |||||||||| |||||||
Db         86 -TGAGCAGCATGGCTTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCT 144

Qy        124 CTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGC 183
              | |||||||||||| | || |||||||||||||||||| |||||| | ||||||||||||
Db        145 CCAGGCAAGTTCAACTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGC 204

Qy        184 CGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTG 243
              | || ||||||||  | || | |||||||||  ||||||| |||||||| ||||| ||||
Db        205 CAGTGACAGCTGAATTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTG 264

Qy        244 ATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGG 303
              |||||||||  ||| ||||| |||||||||||||||| |||||||||   |   ||   || ||
Db        265 ATGTCGACAGTGTCAAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGA 324

Qy        304 CAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCT 363
                |||||||  || || ||| ||||||||| || ||| ||||| |||||||||| ||||
Db        325 ACAACGTGGAGGACCTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCT 384

Qy        364 CCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAGTTTA 423
              |||| || |||||||| |||||||||| |||||||||||||||| ||||||||| | |
Db        385 CCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCA 444

Qy        424 TAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGT 483
              |||||||   |||||||| |||||     |||||| |||| ||||| ||||| |||||||
Db        445 TAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGT 504

Qy        484 ACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGA 543
              ||||||| |||||||||||||| || ||||| ||||||| |  ||||||||||| ||||
Db        505 ACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGA 564

Qy        544 TTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAG 603
              | |||||||||||||||||||||||||||| || |||||||||| ||||||||||||||
Db        565 TCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAG 624

Qy        604 GC------------------------------------------------------- 605
              ||
Db        625 GCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGG 684

Qy        606 ----------------------------------------------------CATAATC 612
                                                                   | ||||
Db        685 ATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATC 744

Qy        613 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 672
              |||||||||| |||||||| |||     |   |||||| |   ||| | || ||||||||||
Db        745 GAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGA 804

Qy        673 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 732
              | || ||||| |  |||||||||||||||||| ||||||| || |||||||| ||||| |
Db        805 ACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTG 864

Qy        733 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 792
              |||||||||||||||||||||||||||| || ||||| ||||||||||||||   ||||
Db        865 AAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCT 924

Qy        793 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 852
              |||||||||||||| |||| |||||||| |||||||| ||||| |||||||||| || |
Db        925 TGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAA 984

Qy        853 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 912
              || ||||||| |||||||||||   |||| |||||||| ||||| || ||||||||||||
Db        985 TACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAG 1044

Qy        913 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 972
              | ||||| |||||||||||||||| ||||| |   | || ||||||| |||||||| ||||| |
Db       1045 TGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACG 1104

Qy        973 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1032
              ||  ||||||||||  ||||| ||||||||||||||   ||||||||||||||||||||
Db       1105 GCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATT 1164

Qy       1033 TGGGCATCAAATATTCGTT---------------------------------------- 1051
              |||||||||||||||||||
Db       1165 TGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGC 1224

Qy       1052 -----------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAATAG 1100
                          ||||||||||||||| |||  |||||||| |||||  ||||||||||||
Db       1225 TGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAATAG 1284

Qy       1101 CTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTT 1160
              |||||||||||| |||||  |||||||||      | | |      ||      ||||
Db       1285 TTTGGCATGTCATCAGGAACACTTAATGCCC------TAACCTCCGCTCTCATTATTTTT 1338

Qy       1161 AATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGAT--TATTTTTAAGTTTTATCC 1218
              | |||| ||||| |||||| ||  ||   ||| || || |  ||  || ||    | |||
Db       1339 ATTTTATTGATTTCAGCAACACTTAACTGTTGCATTAGCTTCTAAGTTGAATCAGTTTCC 1398

Qy       1219 GTAGTTTTGATAAAAGAT 1236
              | |||||| | ||  ||
Db       1399 TTGGTTTTGTTGAAGAAT 1416


RESULT 5
BG616950
LOCUS       BG616950                780 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602615625F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4734247 5',
            mRNA sequence.
ACCESSION   BG616950
VERSION     BG616950.1  GI:13668321
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 780)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1597  row: b  column: 08
            High quality sequence stop: 754.
FEATURES             Location/Qualifiers
     source          1. .780
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4734247"
```

```
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      237 a    147 c    146 g    250 t
ORIGIN

  Query Match            39.6%;  Score 622.8;  DB 10;  Length 780;
  Best Local Similarity   91.5%;  Pred. No. 7.2e-66;
  Matches  703;  Conservative    0;  Mismatches   12;  Indels   53;  Gaps    2;

Qy        850 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTC 60

Qy        910 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACAC 969
              |||||||||||||||||||||||||||||| |||| ||||||||||||||||||||||
Db         61 TAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACAC 120

Qy        970 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATG 180

Qy       1030 ATTTGGGCATCAAATATTCGT---------------------------------------- 1050
              |||||||||||||||||||
Db        181 ATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCT 240

Qy       1051 ------------TTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097
                         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 300

Qy       1098 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 1157
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TAGCTTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTAT 360

Qy       1158 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 1217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TTTAATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATC 420

Qy       1218 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 1277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CGTAGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCT 480

Qy       1278 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAA 1337
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        481 ACTTTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTT-AAAAAAAATTGTAA 539

Qy       1338 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAACT 1397
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        540 AAGTCTAGTTACCTACTTTTTCTTTGATTTTCGACGTTTGACTAGCCATCTCAAGCAAGT 599

Qy       1398 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAAAGATCATCTCACGCTGATCAT 1457
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        600 TTCGACGTTTGACTAGCCATCTCAAGCAAGTTTAATCAATGATCATCTCACGCTGATCAT 659

Qy       1458 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGCTCCAAA 1517
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db        660 TGGATCCTACTCAACAAAAGGAAGGGTGGTCAGAAGTACATTAAAGATTTCTGGTCCAAA 719

Qy       1518 TTTTCAATAAATTTCTTCTTCTCCTTTAAAAAAAAAAAAAAAAAAAAA 1565
              ||||||||||||||| || | | |||  ||||||||||||||||||
Db        720 TTTTCAATAAATTTCTGGTTGTGCCTTAGTAAAAAAAAAAAAAAAAAA 767


RESULT 6
```

```
BG567264
LOCUS       BG567264                 806 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602589745F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723690 5',
            mRNA sequence.
ACCESSION   BG567264
VERSION     BG567264.1  GI:13574917
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 806)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1583  row: j  column: 11
            High quality sequence stop: 772.
FEATURES             Location/Qualifiers
     source          1. .806
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4723690"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      239 a    158 c    162 g    247 t
ORIGIN

  Query Match            39.1%;  Score 615.6;  DB 10;  Length 806;
  Best Local Similarity   90.7%;  Pred. No. 5.2e-65;
  Matches  724;  Conservative    0;  Mismatches   14;  Indels   60;  Gaps    4;

Qy        614 AATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAA 673
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAA 60

Qy        674 TAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGA 733
              ||| |||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGA 120

Qy        734 AACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTT 793
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTT 180

Qy        794 GAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATAT 853
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATAT 240

Qy        854 AGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGT 913
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGT 300

Qy        914 AGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGG 973
              |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
```

```
Db        301 AGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGG 360

Qy        974 CCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTT 1033
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTT 420

Qy       1034 GGGCATCAAATATTCGTTT---------------------------------------- 1052
              |||||||||||||||||||
Db        421 GGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCT 480

Qy       1053 -----------ACATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAATAGC 1101
                         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAATAGC 540

Qy       1102 TTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTTA 1161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TTGGCATGTCATTAGGAATGTTTAATGCCCCTGATTTTATCATTCTGCTTCCGTATTTTA 600

Qy       1162 ATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCCGTA 1221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATTTACTGATTCCAGCAAGACCAAATCATTGTATCAGATTATTTTTAAGTTTTATCCGTA 660

Qy       1222 -GTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGCTACT 1280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db        661 GGTTTTGATAAAAGATTTTCCTATTCCTTGGTTCTGTCAGAGAACCTAATAAGTGTAA-- 718

Qy       1281 TTGCCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAAAATTGTAAAAG 1340
              |  ||||||||||||||||||||||||||||||||||||||||||||||  |    ||||
Db        719 -TTGCATTAAGGCAGACTAGGGTTCATGTCTTTTTACCCTTTAAAAAAATGT----AAAG 773

Qy       1341 TCTAGTTACCTACTTTTT 1358
              |||||||||  |  ||  ||
Db        774 TCTAGTTACTAATTTCTT 791
```

```
RESULT 7
BG618629
LOCUS       BG618629                788 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
            mRNA sequence.
ACCESSION   BG618629
VERSION     BG618629.1  GI:13670000
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: o  column: 15
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767974"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      222 a    157 c    169 g    240 t
ORIGIN
```

```
  Query Match             38.4%;  Score 604.2;  DB 10;  Length 788;
  Best Local Similarity   99.5%;  Pred. No. 1.2e-63;
  Matches  606;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         49 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 108

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        109 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 168

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        169 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 228

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        229 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCAT 288

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        289 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 348

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        349 TGGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 408

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        409 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 468

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        469 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 528

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        529 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATAT 588

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        589 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 648

Qy        601 TAGGCCATA 609
              |||||||||
Db        649 TAGGCCATA 657
```

```
RESULT 8
HSM089979
ID   HSM089979  standard; RNA; EST; 702 BP.
XX
AC   BX496754;
XX
SV   BX496754.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp779I1529_r1  (from clone DKFZp779I1529)
XX
KW   EST; expressed sequence tag.
```

```
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-702
RA   Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by EMBL (European Molecular Biology Laboratories,
CC   Heidelberg/Germany) within the cDNA sequencing consortium of
CC   the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp779I1529) is available at the RZPD in Berlin.
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .702
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp779I1529"
FT                   /clone_lib="779 (synonym: hncc1). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="fetal"
FT                   /tissue_type="liver"
XX
SQ   Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;

  Query Match             37.8%;  Score 594.8;  DB 2;  Length 702;
  Best Local Similarity   99.5%;  Pred. No. 1.8e-62;
  Matches  607;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         42 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 101

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        102 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 161

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        162 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 221

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAG-AAAAAACAAGTCCATTTTTTTGTAAATGCA 239
              ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db        222 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAAACAAGTCCATTTTTTTGTAAATGCA 281

Qy        240 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        282 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 341

Qy        300 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        342 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 401

Qy        360 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        402 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 461

Qy        420 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        462 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 521

Qy        480 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 539
              |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        522 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 581

Qy        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        582 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 641

Qy        600 ATAGGCCATA 609
              ||||||||||
Db        642 ATAGGCCATA 651


RESULT 9
BG566051
LOCUS       BG566051                747 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602582652F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
            mRNA sequence.
ACCESSION   BG566051
VERSION     BG566051.1  GI:13573704
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1549  row: g  column: 13
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4710564"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    157 c    159 g    219 t
ORIGIN

  Query Match             36.9%;  Score 581.2;  DB 10;  Length 747;
  Best Local Similarity   99.2%;  Pred. No. 7.2e-61;
  Matches  605;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db          4 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTG-TCTCT 62

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||| ||
Db         63 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTATAA 122
```

```
Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 182

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCAT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCAT 242

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 302

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 362

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 422

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 482

Qy        481 AGTACCCACTCTATGTTT-TAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 539
              |||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db        483 AGTACCCACTCTATGTTTCTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 542

Qy        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 599
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        543 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 602

Qy        600 ATAGGCCATA 609
              ||||||||||
Db        603 ATAGGCCATA 612
```

RESULT 10
BG618239
LOCUS BG618239 725 bp mRNA linear EST 18-APR-2001
DEFINITION 602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5', mRNA sequence.
ACCESSION BG618239
VERSION BG618239.1 GI:13669610
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1629 row: f column: 17
High quality sequence stop: 723.
FEATURES Location/Qualifiers
source 1. .725
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4767376"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 203 a 153 c 149 g 220 t
ORIGIN

Query Match 36.8%; Score 579.2; DB 10; Length 725;
Best Local Similarity 99.2%; Pred. No. 1.3e-60;
Matches 603; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

```
Qy          1 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         48 AGAAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCT 107

Qy         61 TCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 120
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        108 TCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAA 167

Qy        121 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGC 180
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        168 CCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGA-TGTTCTCTGGC 226

Qy        181 AGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCAT 240
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        227 AGCCGGTAACAGCTGACCTTA-TGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCAT 285

Qy        241 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 CTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGC 345

Qy        301 TGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 360
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 TGGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAG 405

Qy        361 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 CCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAAT 465

Qy        421 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 TTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGA 525

Qy        481 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATAT 540
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        526 AGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATAT 585

Qy        541 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        586 GGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCA 645

Qy        601 TAGGCCAT 608
              ||||||||
Db        646 TAGGCCAT 653
```

RESULT 11
BG616456
LOCUS BG616456 738 bp mRNA linear EST 18-APR-2001
DEFINITION 602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5', mRNA sequence.
ACCESSION BG616456
VERSION BG616456.1 GI:13667827
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1595  row: c  column: 05
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4733500"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      209 a    161 c    153 g    215 t
ORIGIN

  Query Match             35.7%;  Score 561.8;  DB 10;  Length 738;
  Best Local Similarity   99.6%;  Pred. No. 1.5e-58;
  Matches  563;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          45 GTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTA 104
               ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db           1 GTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTA 60

Qy         105 GCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATAT 164
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 GCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATAT 120

Qy         165 GAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCAT 224
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCAT 180

Qy         225 TTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATT 284
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 TTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATT 240

Qy         285 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGAC 344
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGAC 300

Qy         345 ACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATC 404
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 ACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATC 360

Qy         405 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 464
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 420

Qy         465 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA 524
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db         421 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCA 480

Qy         525 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 584
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         481 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 540

Qy         585 TTCTGCTTGTGGTTCATAGGCCATA 609
               |||||||||||||||||||||||||
Db         541 TTCTGCTTGTGGTTCATAGGCCATA 565


RESULT 12
AV646979
LOCUS       AV646979                735 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.
ACCESSION   AV646979
VERSION     AV646979.1  GI:9867993
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCATD06"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      200 a    158 c    172 g    205 t
ORIGIN

  Query Match             35.3%;  Score 555.2;  DB 9;  Length 735;
  Best Local Similarity   98.4%;  Pred. No. 9.6e-58;
  Matches  571;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy          31 GCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGA 90
               ||||||| |||||||||||||||||||||||||||||||||||||||||||||| |||| |
Db           5 GCCTTGCGGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTTCAGA 64

Qy          91 GTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATC 150
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          65 GTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATC 124

Qy         151 TTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGA 210
               |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db         125 TTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGGTGACCTTATTGTGAAGA 184

Qy         211 AAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAA 270
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         185 AAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAA 244
```

```
Qy        271 ATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGC 330
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 ATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGC 304

Qy        331 AGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACT 390
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        305 AGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGGACTATGAACAGTATCACT 364

Qy        391 CACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTA 450
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 CACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTA 424

Qy        451 CAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAA-GGTTTCT 509
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db        425 CAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGGTGTCT 484

Qy        510 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 544

Qy        570 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 609
              |||||||||||||||||| ||| || ||||||||||||
Db        545 TGGATCTCTCCTGCTTTCTGGTTGGGGGTCATAGGCCATA 584
```

RESULT 13
CB157619

```
LOCUS       CB157619                541 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0216807 L17N670205n1 Homo sapiens cDNA clone
            L17N670205n1-12-A11 5', mRNA sequence.
ACCESSION   CB157619
VERSION     CB157619.1  GI:28142751
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 12  row: A  column: 11
            High quality sequence stop: 541.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L17N670205n1-12-A11"
                     /sex="F"
                     /lab_host="Top10F'"
                     /clone_lib="L17N670205n1"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
BASE COUNT      156 a    121 c    107 g    157 t
ORIGIN
```

```
  Query Match            34.1%;  Score 536.2;  DB 14;  Length 541;
  Best Local Similarity  99.4%;  Pred. No. 2.2e-55;
  Matches  538;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy         22 AGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCG 81
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCG 60

Qy         82 CGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTC 141
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTC 120

Qy        142 TACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTA 201
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTA 180

Qy        202 TTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAG 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAG 240

Qy        262 CCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTA 321
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        241 CCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTA 300

Qy        322 TTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAAC 381
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAAC 360

Qy        382 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTG 441
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTG 420

Qy        442 ATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAA 501
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAA 480

Qy        502 AGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATG 561
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db        481 AGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATG 540

Qy        562 C 562
              |
Db        541 C 541
```

RESULT 14
BG618813

```
LOCUS       BG618813                750 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
            mRNA sequence.
ACCESSION   BG618813
VERSION     BG618813.1  GI:13670184
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: k  column: 22
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
```

```
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767885"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      219 a    146 c    171 g    213 t      1 others
ORIGIN

  Query Match             33.8%;  Score 532.2;  DB 10;  Length 750;
  Best Local Similarity   98.0%;  Pred. No. 5.4e-55;
  Matches  581;  Conservative    0;  Mismatches    8;  Indels    4;  Gaps    4;

Qy          9 GCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAG 68
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 GCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAG 61

Qy         69 CAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGG 128
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db         62 CAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGG 121

Qy        129 CAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTA 188
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 CAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTA 181

Qy        189 ACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTC 248
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 ACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTC 241

Qy        249 GACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGAC 308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGAC 301

Qy        309 GTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCA 368
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        302 GTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGACCCCGAGCCTCCGCA 361

Qy        369 TCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACT 428
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 TCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACT 421

Qy        429 GAGAGGCATCCTGATATGCTT-ACAAAAATCCACATTGG-ATCCTCATTTGAGAAGTACC 486
              |||||||||||||||||||| |||||||||||||||| ||||||| |||||||| |
Db        422 GAGAGGCATCCTGATATGCTTAACAAAAATCCACATTGGCATCCTCATTCGAGAAGTAAC 481

Qy        487 CACTC-TATGTTTTAAAGGTTTCT-GGAAAAGAACAAACAGCCAAAAATGCCATATGGAT 544
              || || |||||||||||||| || |||||||||||| ||| ||||||||||||||||||
Db        482 CAGTCGTATGTTTTAAAGGTTCCTGGGAAAAGAACAAGCAGGCAAAAATGCCATATGGAT 541

Qy        545 TGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGT 597
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGT 594


RESULT 15
BG568240
LOCUS       BG568240                691 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602587149F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715797 5',
            mRNA sequence.
ACCESSION   BG568240
VERSION     BG568240.1  GI:13575893
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1563  row: a  column: 14
            High quality sequence stop: 691.
FEATURES             Location/Qualifiers
     source          1. .691
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4715797"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      204 a    143 c    144 g    200 t
ORIGIN

  Query Match             33.3%;  Score 523.8;  DB 10;  Length 691;
  Best Local Similarity   98.5%;  Pred. No. 5.8e-54;
  Matches  539;  Conservative    0;  Mismatches    7;  Indels    1;  Gaps    1;

Qy          3 AAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTC 62
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          4 AAAATTGCTGTTGGGATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTC 63

Qy         63 TGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACC 122
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db         64 TGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACC 123

Qy        123 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAG 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        124 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAG 183

Qy        183 CCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCT 242
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db        184 CCGGTAACAGCTGACCTTA-TGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCT 242

Qy        243 GATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTG 302
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 GATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTG 302

Qy        303 GCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCC 362
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 GCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCC 362

Qy        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 422

Qy        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 482
```

```
Qy        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 542
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        483 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATAA 542

Qy        543 ATTGACT 549
              | | | |
Db        543 ACTCAAT 549
```

Search completed: January 1, 2004, 15:37:43
Job time : 3649.93 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:38:01 ; Search time 141.462 Seconds
(without alignments)
2423.606 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 2215726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135747_24935/app_query.fasta_1.2958
-DB=A_Geneseq_19Jun03 -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_117_@runat_31122003_135747_24935 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1911	98.9	360	21	AAB11457	Human brain carbox
2	1857.5	96.1	386	24	ABG72095	Human protease pro
3	1852.5	95.9	386	24	ABG72093	Human protease pro
4	1835	95.0	423	17	AAR90293	Human plasma carbo
5	1835	95.0	423	18	AAW14733	Human plasma carbo
6	1835	95.0	423	20	AAW92270	Human plasma carbo
7	1830	94.7	423	14	AAR36273	Human plasma carbo
8	1780	92.1	338	21	AAB11458	Human brain carbox
9	1320	68.3	246	21	AAB11459	Human brain carbox
10	839	43.4	211	21	AAB58129	Lung cancer associ
11	641	33.2	417	24	ABB99688	Amino acid sequenc
12	640.5	33.2	404	16	AAR75131	Porcine Tyr-His-Me
13	639.5	33.1	415	17	AAW06172	Human pancreatic c
14	639.5	33.1	415	19	AAW74476	Human pancreatic c
15	636	32.9	324	22	AAG66562	Bothops jararaca c
16	635	32.9	417	23	AAU84325	Protein CPA3 diffe
17	634	32.8	667	24	ABP97414	Anti-human seminal
18	632	32.7	424	17	AAW06175	ProHCPB with PelB
19	623	32.2	716	18	AAW41414	PreproHCPB-linker-
20	613.5	31.8	434	24	ABB99687	Amino acid sequenc
21	613.5	31.8	437	22	AAE00502	Human carboxypepti
22	613.5	31.8	437	23	AAE19175	Human protease, PR
23	604	31.3	307	17	AAW00602	Rat carboxypeptida
24	576	29.8	306	16	AAR75132	Porcine carboxypep
25	576	29.8	306	22	AAU04477	Porcine carboxypep
26	564.5	29.2	354	22	AAG66560	Human secreted met
27	564.5	29.2	354	22	AAG66566	Human secreted met
28	564.5	29.2	362	22	AAB74682	Human protease and
29	564.5	29.2	374	22	AAG66547	Human secreted met
30	564.5	29.2	374	22	AAG66565	Human secreted met
31	564.5	29.2	374	23	AAU82703	Amino acid sequenc
32	564.5	29.2	374	24	ABJ37888	NOVX protein seque
33	564.5	29.2	374	24	ABG74658	Human zinc-binding
34	563	29.1	444	23	AAE24240	Human 23566 (carbo
35	561	29.0	329	17	AAW06174	Mature HCPB with P
36	561	29.0	349	17	AAW06173	mature HCPB-(His)6
37	560	29.0	613	18	AAW41412	Humanised Fd-mutan
38	560	29.0	613	19	AAW74481	Human pancreatic c
39	554	28.7	349	17	AAW06181	Modified HCPB (D25
40	554	28.7	349	18	AAW13749	Carboxypeptidase B
41	553	28.6	349	17	AAW06182	Modified HCPB (D25
42	553	28.6	349	18	AAW13750	Carboxypeptidase B
43	550	28.5	349	18	AAW13763	Carboxypeptidase B
44	548.5	28.4	374	24	ABJ37887	NOVX protein seque
45	548	28.4	349	18	AAW13752	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAB11457
ID AAB11457 standard; Protein; 360 AA.
XX
AC AAB11457;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX

PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
DR N-PSDB; AAC81962.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 1; Page 68-71; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 360 AA;

Alignment Scores:

Pred. No.:	6.96e-198	Length:	360
Score:	1911.00	Matches:	360
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	98.91%	Indels:	0
DB:	21	Gaps:	0

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB11457 (1-360)

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1021 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
```

RESULT 2
ABG72095
ID ABG72095 standard; Protein; 386 AA.
XX
AC ABG72095;
XX
DT 14-FEB-2003 (first entry)
XX
DE Human protease protein, variation #2.
XX
KW Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 169
FT /note= "This residue is the minor variant of a single
FT nucleotide polymorphism. The major variant is an
FT Ala residue"
XX
PN US2002137179-A1.
XX
PD 26-SEP-2002.
XX
PF 21-MAR-2001; 2001US-0813133.
XX
PR 21-MAR-2001; 2001US-0813133.

```
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Page -; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #2, the gene for which is located on chromosome 13.
CC   Note: This sequence is not shown in the specification but has been
CC   created by the indexer from information given in figure 3.
XX
SQ   Sequence   386 AA;

Alignment Scores:
Pred. No.:                4.5e-192        Length:          386
Score:                    1857.50         Matches:         355
Percent Similarity:       94.43%          Conservative:    1
Best Local Similarity:    94.16%          Mismatches:      4
Query Match:              96.14%          Indels:          18
DB:                       24              Gaps:            1

US-09-980-881A-1_COPY_18_1097 (1-1080) x ABG72095 (1-386)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              ||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1021 ATCAAATATTCGTT---------------------------------------------- 1034
              ||||||||||||
Db        341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360

Qy       1035 ------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
                     ||||||||||||||||||||||||||||||||||||||||||||
Db        361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSerLys 377

RESULT 3
ABG72093
ID   ABG72093 standard; Protein; 386 AA.
XX
AC   ABG72093;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #1.
XX
KW   Human; enzyme; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
```

```
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 169
FT                   /note= "This residue is the major variant of a single
FT                          nucleotide polymorphism. The minor variant is an
FT                          Thr residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Fig 2; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #1, the gene for which is located on chromosome 13.
XX
SQ   Sequence   386 AA;

Alignment Scores:
Pred. No.:              1.57e-191       Length:         386
Score:                  1852.50         Matches:        354
Percent Similarity:     94.16%          Conservative:   1
Best Local Similarity:  93.90%          Mismatches:     5
Query Match:            95.89%          Indels:         18
DB:                     24              Gaps:           1

US-09-980-881A-1_COPY_18_1097 (1-1080) x ABG72093 (1-386)
```

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              ||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1021 ATCAAATATTCGTT---------------------------------------------- 1034
              ||||||||||||
Db        341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360

Qy       1035 ------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
```

```
                     ||||||||||||||||||||||||||||||||||||||||||||
Db       361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSerLys 377

RESULT 4
AAR90293
ID   AAR90293 standard; Protein; 423 AA.
XX
AC   AAR90293;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active-site     181
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    183
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    256..257
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     310
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    312
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     369
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    370..371
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    394
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   N-PSDB; AAT11671.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

Alignment Scores:
Pred. No.:              1.29e-189      Length:          423
Score:                  1835.00        Matches:         356
Percent Similarity:     86.23%         Conservative:    1
Best Local Similarity:  85.99%         Mismatches:      3
Query Match:            94.98%         Indels:          55
DB:                     17             Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAR90293 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280
```

```
Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
```

```
RESULT 5
AAW14733
ID   AAW14733 standard; Protein; 423 AA.
XX
AC   AAW14733;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "Signal peptide"
FT   Protein         23..423
FT                   /note= "Mature PCPB"
FT   Cleavage-site   114
FT                   /note= "Potential clip site for activation of PCPB
FT                          as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    183
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     184
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    256
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    257
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     310
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    312
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     369
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    370
FT                   /note= "Expected to be involved in substrate
FT                          binding, thought to determine specificity
FT                          of PCPB as a carboxypeptidase B"
FT   Binding-site    371
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    394
FT                   /note= "Expected to be involved in substrate binding"
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   N-PSDB; AAT62846.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence represents human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;
```

```
Alignment Scores:
Pred. No.:              1.29e-189      Length:          423
Score:                  1835.00        Matches:         356
Percent Similarity:     86.23%         Conservative:    1
Best Local Similarity:  85.99%         Mismatches:      3
Query Match:            94.98%         Indels:          55
DB:                     18             Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAW14733 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
```

```
Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT-------------------------------------------------------TAC 1037
              |||                                                         |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 6
AAW92270
ID   AAW92270 standard; Protein; 423 AA.
XX
AC   AAW92270;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW   polymorphism detection; thrombotic disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "signal peptide"
FT   Protein         23..423
FT                   /note= "mature PCPBthr147"
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   N-PSDB; AAV74302.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Disclosure; Page 25-26; 35pp; English.
XX
CC   This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence   423 AA;

Alignment Scores:
Pred. No.:              1.29e-189      Length:          423
Score:                  1835.00        Matches:         356
Percent Similarity:     86.23%         Conservative:    1
Best Local Similarity:  85.99%         Mismatches:      3
Query Match:            94.98%         Indels:          55
DB:                     20             Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAW92270 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy         301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy         361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy         421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy         481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy         541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy         591 ------------------------------------------------------------ 591

Db         201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy         592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db         221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy         610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy         670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy         730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy         790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy         850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy         910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
                ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db         341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy         970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy        1030 TCGTT-----------------------------------------------------TAC 1037
                |||                                                     |||
Db         381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy        1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
                |||||||||||||||||||||||||||||||||||||||||||
Db         401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 7
AAR36273
ID   AAR36273 standard; Protein; 423 AA.
XX
AC   AAR36273;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= signal_peptide
FT   Cleavage-site   114
FT                   /note= "putative site of cleavage by trypsin to
FT                         activate PCPB as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "expected to be involved in catalytic
FT                         activity"
FT   Binding-site    183
FT                   /note= "expected to be involved in substrate
FT                         binding"
FT   Active-site     184
FT                   /note= "expected to be involved in catalytic
FT                         activity"
FT   Binding-site    256..257
FT                   /note= "expected to be involved in substrate
FT                         binding"
FT   Active-site     310
FT                   /note= "expected to be involved in catalytic
FT                         activity"
FT   Binding-site    312
FT                   /note= "expected to be involved in substrate
FT                         binding"
FT   Active-site     369
FT                   /note= "expected to be involved in catalytic
FT                         activity"
FT   Binding-site    370..371
FT                   /note= "expected to be involved in substrate
FT                         binding; Asp 370 is the residue which
FT                         determines specificity of the PCPB as
FT                         a carboxypeptidase B"
FT   Binding-site    394
FT                   /note= "expected to be involved in substrate
FT                         binding"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   N-PSDB; AAQ41001.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Claim 2; Fig 4; 40pp; English.
XX
CC   This amino acid sequence was deduced from a clone isolated from a
CC   human liver cDNA library. There is about 40% sequence identity
CC   between the deduced preproPCPB amino acid sequence and pro-human
CC   mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC   carboxypeptidase B. Human PCPB has the same substrate binding sites
CC   as, and shares 6 cysteine residues which form 3 intramolecular
CC   disulphide bonds with, bovine and rat carboxypeptidase B. The
CC   presence of the same amino acid (Asp at position 348 of the mature
```

```
CC   protein) at the region in carboxypeptidases that determines
CC   substrate specificity, suggests that PCPB represents a plasma-
CC   derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC   tPA of plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence   423 AA;

Alignment Scores:
Pred. No.:              4.49e-189       Length:         423
Score:                  1830.00         Matches:        355
Percent Similarity:     85.99%          Conservative:   1
Best Local Similarity:  85.75%          Mismatches:     4
Query Match:            94.72%          Indels:         55
DB:                     14              Gaps:           2

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAR36273 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaCysSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 8
AAB11458
ID   AAB11458 standard; Protein; 338 AA.
XX
AC   AAB11458;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 71-73; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
```

```
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   338 AA;

Alignment Scores:
Pred. No.:              1.05e-183       Length:         338
Score:                  1780.00         Matches:        335
Percent Similarity:     99.41%          Conservative:   1
Best Local Similarity:  99.11%          Mismatches:     2
Query Match:            92.13%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB11458 (1-338)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 606
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180

Qy        607 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200

Qy        667 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220

Qy        727 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240

Qy        787 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260

Qy        847 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280

Qy        907 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 966
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300

Qy        967 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320

Qy       1027 TATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 TyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338

RESULT 9
AAB11459
ID   AAB11459 standard; Protein; 246 AA.
XX
AC   AAB11459;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 74-75; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   246 AA;

Alignment Scores:
Pred. No.:              6.28e-134       Length:         246
Score:                  1320.00         Matches:        243
Percent Similarity:     99.19%          Conservative:   1
Best Local Similarity:  98.78%          Mismatches:     2
Query Match:            68.32%          Indels:         0
DB:                     21              Gaps:           0

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB11459 (1-246)

Qy        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
```

```
Qy        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

Qy        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 522
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60

Qy        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80

Qy        583 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 642
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100

Qy        643 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 702
              ||||||||||||||    ||||||  ||||||||||||||||||||||||||||||||
Db        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120

Qy        703 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 762
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140

Qy        763 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 822
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160

Qy        823 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 882
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180

Qy        883 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 942
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200

Qy        943 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1002
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220

Qy       1003 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTT 1062
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 TrpIleTyrAspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeu 240

Qy       1063 TTGCCGCTGTCTCTAAAA 1080
              ||||||||||||||||||
Db        241 LeuProLeuSerLeuLys 246

RESULT 10
AAB58129
ID   AAB58129 standard; Protein; 211 AA.
XX
AC   AAB58129;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18005.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 944-945; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   211 AA;

Alignment Scores:
Pred. No.:              7.62e-82        Length:         211
Score:                  839.00          Matches:        164
Percent Similarity:     82.59%          Conservative:   2
Best Local Similarity:  81.59%          Mismatches:     9
Query Match:            43.43%          Indels:         27
DB:                     21              Gaps:           3

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAB58129 (1-211)

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------------GGC 588
              |||||||||          |||  |||   :::
Db          3 TrpIleSer---MetLeuCysArgTrpLeuMetTrpMetValMetAsnTyrSerTrpLys 21

Qy        589 CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACA 648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         22 LysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThr 41

Qy        649 GACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCA 708
              |||||||||   ||||||  ||||||||||||||||||||||||||||||||||||||||
Db         42 AspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSer 61

Qy        709 TGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCT 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAla 81

Qy        769 AGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCC 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 SerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSer 101

Qy        829 CAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTG 888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 GlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 121

Qy        889 TCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTAT 948
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
```

```
Db        122 SerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyr 141

Qy        949 ACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATC 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 ThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIle 161

Qy       1009 TATGATTTGGGCATCAAATATTCGTT---------------------------------- 1034
              ||||||||||||||||||||||||
Db        162 TyrAspLeuGlyIleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGly 181

Qy       1035 ------------------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCT 1076
                               ||||||||||||||||||||||||||||||||||||||||||
Db        182 PheLeuLeuProGluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSer 201

Qy       1077 AAA 1079
              |||
Db        202 Lys 202

RESULT 11
ABB99688
ID   ABB99688 standard; Protein; 417 AA.
XX
AC   ABB99688;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Amino acid sequence of human zinc carboxypeptidase B precursor.
XX
KW   Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW   chronic obstructive pulmonary disease; COPD; haematological disorder;
KW   zinc carboxypeptidase B.
XX
OS   Homo sapiens.
XX
PN   WO200295018-A1.
XX
PD   28-NOV-2002.
XX
PF   16-MAY-2002; 2002WO-EP05392.
XX
PR   22-MAY-2001; 2001US-292329P.
XX
PA   (FARB ) BAYER AG.
XX
PI   Smolyar A;
XX
DR   WPI; 2003-120801/11.
XX
PT   New polynucleotide encoding a zinc carboxypeptidase B-like protein
PT   polypeptide, useful for treating diseases associated with
PT   carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT   haematological disorder  -
XX
PS   Disclosure; Fig 3; 113pp; English.
XX
CC   The present sequence represents a zinc carboxypeptidase B precursor.
CC   The specification describes a human zinc carboxypeptidase B-like
CC   protein. The zinc carboxypeptidase B-like protein is used for treating
CC   diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
CC   disease (COPD) or haematological disorders. The zinc carboxypeptidase
CC   B-like protein may also be used to identify compounds which may act
CC   as activators or inhibitors at the enzyme's active site, to raise
CC   specific antibodies which can block the enzyme and effectively reduce
CC   its activity, as a bait protein in a two-hybrid or three-hybrid assay
CC   to identify other proteins which bind to or interact with the human
CC   zinc carboxypeptidase-like protein polypeptide and modulate its
CC   activity, and for immunization of mammals.
XX
SQ   Sequence   417 AA;

Alignment Scores:
Pred. No.:              2.98e-60        Length:         417
Score:                  641.00          Matches:        140
Percent Similarity:     51.82%          Conservative:   73
Best Local Similarity:  34.06%          Mismatches:     137
Query Match:            33.18%          Indels:         62
DB:                     24              Gaps:           8

US-09-980-881A-1_COPY_18_1097 (1-1080) x ABB99688 (1-417)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              :::   :::|||  |||||||||  :::  |||        |||
Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         67 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
                 |||:::      :::|||                                :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 183
              ::::::   |||   :::|||  :::|||      |||:::|||   :::
Db         41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        184 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 243
              |||                |||  |||  |||  |||  |||  |||:::
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy        244 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 303
                 |||      :::  :::         ||||||:::::::::::::   ::::::
Db         81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy        304 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::   |||     :::                ||||||:::          ||||||
Db        101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||:::          |||  :::|||  :::  :::          :::|||
Db        118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
                 :::::::::::      ||||||::::::||||||         :::|||:::|||
Db        138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              ||||||   ||||||          |||   ||||||::::::|||||||||   |||
Db        158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
              ||||||||||||||||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyArgGluIleGlnValThrGluLeuLeuAsnLysLeuAspPheTyrValLeuPro 216

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 :::|||   |||||||||
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||   :::      :::   |||||||||||||||   |||   ||||||
Db        237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                  |||||||||   |||||||||   :::   |||   |||||||||||||||
Db        256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                  |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              ||||||||||||:::::::::|||||||||||||||   :::::::::|||||||||
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              |||                 :::::::::   ||||||:::   :::|||         |||
```

```
Db        316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
              :::   :::      :::         |||:::||||||:::|||   |||:::   |||
Db        336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              :::|||   |||   ||||||   |||||||||   ||||||   |||||:::||||||
Db        355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

Qy       1033 TTTACATC---------------------------------------------------- 1040
              ||||||
Db        375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              :::   ||||||   |||   |||   |||:::
Db        395 ArgAlaThrCysGluGluThrPheLeuAlaIle 405

RESULT 12
AAR75131
ID   AAR75131 standard; Protein; 404 AA.
XX
AC   AAR75131;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine Tyr-His-Met Procarboxypeptidase B.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB.
XX
OS   Sus scrofa.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   N-PSDB; AAQ90600.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Example 1; Page 16-17; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence   404 AA;

Alignment Scores:
Pred. No.:              3.33e-60        Length:         404
Score:                  640.50          Matches:        136
Percent Similarity:     54.29%          Conservative:   60

Best Local Similarity:  37.67%          Mismatches:     122
Query Match:            33.15%          Indels:         43
DB:                     16              Gaps:           6

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAR75131 (1-404)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||:::      :::|||                  :::   :::|||
Db         10 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerLeuLeu 29

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
                  |||   :::|||   :::|||      |||:::|||   :::         |||
Db         30 HisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGlnIle 49

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                           |||   |||   |||   |||   |||:::      |||:::
Db         50 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 69

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                 |||      :::  :::           ||||||:::   ::::::      ::::::
Db         70 PheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerValLeu 89

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
              :::   |||      :::                |||   :::         ||||||:::
Db         90 GluAlaGlnPheAspSerArgVal---------ArgThrThrGlyHisSerTyrGluLys 106

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
              |||::::::          |||   :::|||   :::   :::|||      :::||||||
Db        107 TyrAsnAsnTrpGluThrIleGluAlaTrpThrLysGlnValThrSerGluAsnProAsp 126

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
              ::::::::::::          ||||||::::::|||          :::|||:::||||||
Db        127 LeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeuLys 146

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
              |||   ||||||            |||   ||||||::::::|||||||||   ||||||
Db        147 Val---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHisAla 165

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
              |||||||||||||||   |||||||||   ||||||:::
Db        166 ArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThrTyr 185

Qy        585 ------------------------------------------------------------ 585

Db        186 GlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuProVal 205

Qy        586 ------------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGT 615
                         |||:::                 ||||||||||||||||||   |||
Db        206 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 225

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||      |||          ||||||||||||||   |||   ||||||
Db        226 SerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---AspAla 244

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
              ||||||      |||||||||:::      |||   ||||||||||||||
Db        245 GlyTrpCysThrThrGlyAlaSerThrAspProCysAspGluThrTyrCysGlySerAla 264

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
              |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        265 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSerSer 284

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              ||||||||||||:::::::::|||||||||||||||   |||::::::||||||||||||
Db        285 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 304

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
                           :::::::::   ||||||:::   :::|||      ||||||:::
Db        305 AspTyrLysLeuProGluAsnAsnAlaGluLeuAsnAsnLeuAlaLysAlaAlaValLys 324

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
              :::      |||          |||:::||||||:::|||   |||:::   |||:::
Db        325 GluLeu---AlaThrLeuTyrGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 343
```

```
Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||   |||   ||||||   |||||||||   ||||||   ||||||||||||||||||
Db        344 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrSerPhe 363

Qy       1036 ACA 1038
              |||
Db        364 Thr 364

RESULT 13
AAW06172
ID   AAW06172 standard; Protein; 415 AA.
XX
AC   AAW06172;
XX
DT   12-FEB-1997  (first entry)
XX
DE   Human pancreatic carboxypeptidase B.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..13
FT                   /note= "partial enzyme secretion leader sequence;
FT                         (pre-sequence)"
FT   Peptide         14..108
FT                   /label= pro-sequence
FT   Protein         109..415
FT                   /label= mature_enzyme
FT   Misc-difference 243
FT                   /note= "not observed in other published HPCB
FT                         sequences"
FT   Active-site     361
FT                   /note= "Asp253 characteristic of B type specificity"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   N-PSDB; AAT42494.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 14; Page 131-132; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is that of the native prepro-HCPB.
XX
SQ   Sequence   415 AA;

Alignment Scores:
Pred. No.:              4.32e-60        Length:         415
Score:                  639.50          Matches:        140
Percent Similarity:     51.72%          Conservative:   71
Best Local Similarity:  34.31%          Mismatches:     136
Query Match:            33.10%          Indels:         62
DB:                     17              Gaps:           8

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAW06172 (1-415)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 72
              |||   |||||||||   :::   |||           |||           |||:::
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         73 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 132
                 :::|||                       :::   :::::::::
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        133 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 192
              |||   :::|||   :::|||      |||:::|||   :::         |||
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        193 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 252
                    |||   |||   |||   |||   |||         |||:::      |||
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        253 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 312
                  :::   :::          ||||||:::::::::::   ::::::::::::
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        313 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
              |||       :::                 ||||||:::       ||||||:::|||:::
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                       |||   :::|||   :::   :::         :::|||   ::::::
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
              ::::::     ||||||:::::||||||         :::|||:::|||||||||
Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
              ||||||            |||   |||||:::::|||||||||   ||||||||||||
Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 585
              ||||||||||||||||||||   |||||:::
Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        585 ------------------------------------------------------------ 585

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        586 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 621
                    |||:::                 :::|||   |||||||||   ||||||
Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                  :::      :::   ||||||||||||||   |||   ||||||          |||
Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
              ||||||   |||||||||   :::   |||   ||||||||||||||          |||
Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
              ||||||   |||   ||||||:::|||   |||:::|||      ::::::   ||||||
Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
```

```
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921

Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981

Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1040

Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1041 ----------------------------------------------------AAACCCACC 1049

Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073

Db        396 CysGluGluThrPheLeuAlaIle 403

RESULT 14
AAW74476
ID   AAW74476 standard; Protein; 415 AA.
XX
AC   AAW74476;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Human pancreatic carboxypeptidase protein sequence.
XX
KW   Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW   prodrug therapy.
XX
OS   Homo sapiens.
XX
PN   WO9835988-A1.
XX
PD   20-AUG-1998.
XX
PF   10-FEB-1998;   98WO-GB00415.
XX
PR   29-OCT-1997;   97GB-0022727.
PR   14-FEB-1997;   97GB-0003104.
PR   18-OCT-1997;   97GB-0022003.
XX
PA   (ZENE ) ZENEPA   (ZENE ) ZENECA LTD.
XX
PI   Edge MD;
XX
DR   WPI; 1998-467168/40.
DR   N-PSDB; AAV41795.
XX
PT   New modified pro-domain of carboxy-peptidase B - enhances expression
PT   of co-expressed proteins for production of recombinant
PT   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT   enzyme prodrug therapy
XX
PS   Example 1; Page 54-55; 83pp; English.
XX
CC   The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC   modified pro-domain of CPB on a separate gene to enhance recombinant
CC   expression.  This process can be used to produce recombinant CPB in
CC   eukaryotic cells, or fusions of CPB with antibody chains.  CPB is used in
CC   insulin production and protein sequencing, while its fusions with
CC   antibodies are useful in antibody-directed enzyme prodrug therapy.  The
CC   modified pro-domain provide increased yields of recombinant CPB, possibly
CC   by protecting the C-terminus against enzymatic degradation or by
CC   increasing intracellular trafficking.
XX
SQ   Sequence   415 AA;
```

```
Alignment Scores:
Pred. No.:              4.32e-60      Length:          415
Score:                  639.50        Matches:         140
Percent Similarity:     51.72%        Conservative:    71
Best Local Similarity:  34.31%        Mismatches:      136
Query Match:            33.10%        Indels:          62
DB:                     19            Gaps:            8

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAW74476 (1-415)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 72

Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         73 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 132

Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        133 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 192

Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        193 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 252

Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        253 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 312

Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        313 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372

Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432

Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492

Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552

Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------------- 585

Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        585 ------------------------------------------------------------ 585

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        586 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 621

Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681

Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741

Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801

Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
```

```
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921

Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981

Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1040

Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1041 ----------------------------------------------------AAACCCACC 1049

Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073

Db        396 CysGluGluThrPheLeuAlaIle 403

RESULT 15
AAG66562
ID   AAG66562 standard; Protein; 324 AA.
XX
AC   AAG66562;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Bothops jararaca carboxypeptidase homologue protein.
XX
KW   secreted metallocarboxypeptidase; immunomodulatory;
KW   antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
KW   osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnerary;
KW   antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
KW   antirheumatic; cytostatic; antiallergic; haemostatic; infection;
KW   cancer.
XX
OS   Bothops jararaca.
XX
PN   WO200157265-A1.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03783.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   29-SEP-2000; 2000US-0676135.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Boyle BJ,  Mize NK,  Arterburn MC,  Yeung G,  Tang YT,  Liu C;
PI   Drmanac RT,  Wang M,  Chen L,  Yang Y;
XX
DR   WPI; 2001-465705/50.
XX
PT   Novel metallocarboxy peptidase-like polypeptide is useful in treating
PT   neurodegenerative, immune, etc. disorders  -
XX
PS   Example 3; Fig 1; 146pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino
CC   acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
CC   354, 165, 374, 354, or 41 amino acids, fully defined in the
CC   specification, or its translated protein coding portion, mature protein
CC   coding portion, extracellular portion or active domain. The polypeptide
CC   is a human secreted metallocarboxypeptidase-like polypeptide. It
CC   is useful for treating mammalian diseases including:
CC   neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
CC   osteoporosis, osteoarthritis, disorders linked to reduced tissue
CC   growth, Huntington's disease, amylotrophic lateral sclerosis, stroke,
CC   reperfusion, wound healing, infectious diseases, immune and autoimmune
CC   disorders such as severe combined immunodeficiency, rheumatoid
CC   arthritis, Guillain-Barre syndrome, insulin dependent diabetes
CC   mellitus, graft-versus-host disease and allergy, cancer, and
CC   haemophilia. The polypeptide is also useful in medical imaging, as an
CC   immunogen to generate antibodies, as a nutritional source or
CC   supplement, and as a contraceptive. The present sequence is
CC   a carboxypeptidase homologue protein of Bothops jararaca.
XX
SQ   Sequence   324 AA;

Alignment Scores:
Pred. No.:              9.21e-60        Length:         324
Score:                  636.00          Matches:        132
Percent Similarity:     56.23%          Conservative:   53
Best Local Similarity:  40.12%          Mismatches:     84
Query Match:            32.92%          Indels:         61
DB:                     22              Gaps:           6

US-09-980-881A-1_COPY_18_1097 (1-1080) x AAG66562 (1-324)

Qy        250 TTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAA 309

Db          2 LeuGlnGlnSerGlyLeuAsnTyrGluIleLeuIleAspAsnLeuGlnAlaValLeuAsp 21

Qy        310 CAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTAT 369

Db         22 ArgGlnLeuAspAsnHisAla---------ArgThrAlaGlyTyrAsnTyrGluLysTyr 38

Qy        370 CACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATG 429

Db         39 AsnSerTrpGluLysIleAspAlaTrpThrAlaAspIleAlaAsnGluAsnProSerLeu 58

Qy        430 CTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTT 489

Db         59 ValSerArgLeuGlnIleGlyThrThrPheGluGlyArgProMetProLeuLeuLysVal 78

Qy        490 TCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGA 549

Db         79 ---GlyLysProGlyValAsnLysLysAlaIlePheIleAspCysGlyPheHisAlaArg 97

Qy        550 GAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585

Db         98 GluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGly 117

Qy        585 ------------------------------------------------------------ 585

Db        118 LysGluThrIleMetThrGlnLeuLeuAsnLysLeuAspPheTyrIleLeuProValLeu 137

Qy        586 --------------------------GGCCATAATCGAATGTGGAGAAAGAACCGTTCT 618

Db        138 AsnIleAspGlyTyrValTyrSerTrpLysGlnSerArgMetTrpArgLysThrArgSer 157

Qy        619 TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACAC 678

Db        158 ValAsnAlaGlySerThrCysIleGlyThrAspProAsnArgAsnPhe---AspAlaAla 176

Qy        679 TGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCT 738

Db        177 TrpCysSerValGlyAlaSerArgAsnProCysSerGluThrTyrCysGlySerLysPro 196

Qy        739 GAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATT 798

Db        197 GluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgArgAsnArgSerIleIle 216

Qy        799 AAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACA 858

Db        217 GlnAlaTyrLeuThrIleHisSerTyrSerGlnMetLeuLeuTyrProTyrSerTyrThr 236

Qy        859 CGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCT 918

Db        237 TyrAspLeuThrSerAsnAsnLysLysLeuAsnSerIleAlaLysGluAlaIleArgGlu 256
```

```
Qy        919 ATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATAC 978
              :::   |||           |||  ||||||:::|||   |||:::   |||:::|||
Db        257 Leu---LysValLeuPheGlyThrGluTyrThrTyrGlyProGlyAlaAlaThrIleTyr 275

Qy        979 CTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
                |||   ||||||   |||||||||   ||||||   ||||||||||||:::||||||
Db        276 ProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrAlaPheThr 295

Qy       1039 TC-------------------------------------------------AAACCC 1046
                                                                 ||||||
Db        296 PheGluLeuArgAspLysGlyArgTyrGlyPheAlaLeuProGluSerGlnIleLysPro 315

Qy       1047 ACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              ||||||   |||         ||||||
Db        316 ThrCysGluGluThrMetIleAlaVal 324
```

Search completed: January 1, 2004, 18:52:58
Job time : 160.462 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:40:51 ; Search time 29.1067 Seconds
(without alignments)
3139.880 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 657434

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135750_24983/app_query.fasta_1.2958
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_31_@runat_31122003_135750_24983 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1852.5	95.9	386	4	US-09-813-133A-2	Sequence 2, Appli
2	1835	95.0	423	1	US-07-649-591B-3	Sequence 3, Appli
3	1835	95.0	423	1	US-08-277-540-3	Sequence 3, Appli
4	1835	95.0	423	1	US-08-430-787A-3	Sequence 3, Appli
5	1835	95.0	423	2	US-08-869-057-2	Sequence 2, Appli
6	1824	94.4	423	4	US-09-813-133A-4	Sequence 4, Appli
7	640.5	33.2	404	1	US-08-696-139-2	Sequence 2, Appli
8	639.5	33.1	415	2	US-08-860-882A-57	Sequence 57, Appl
9	639.5	33.1	415	4	US-09-011-769A-39	Sequence 39, Appl
10	638	33.0	417	1	US-07-649-591B-7	Sequence 7, Appli
11	638	33.0	417	1	US-08-277-540-7	Sequence 7, Appli
12	638	33.0	417	1	US-08-430-787A-7	Sequence 7, Appli
13	635	32.9	417	1	US-07-649-591B-6	Sequence 6, Appli
14	635	32.9	417	1	US-08-277-540-6	Sequence 6, Appli
15	635	32.9	417	1	US-08-430-787A-6	Sequence 6, Appli
16	632	32.7	424	4	US-09-011-769A-56	Sequence 56, Appl
17	623	32.2	716	3	US-09-171-945-125	Sequence 125, App
18	613.5	31.8	437	4	US-09-675-305-10	Sequence 10, Appl
19	604	31.3	307	2	US-08-782-760-6	Sequence 6, Appli
20	604	31.3	307	5	PCT-US96-00995-6	Sequence 6, Appli
21	603.5	31.2	396	1	US-07-649-591B-4	Sequence 4, Appli
22	603.5	31.2	396	1	US-08-277-540-4	Sequence 4, Appli
23	603.5	31.2	396	1	US-08-430-787A-4	Sequence 4, Appli
24	576	29.8	306	1	US-08-696-139-4	Sequence 4, Appli
25	561	29.0	329	4	US-09-011-769A-51	Sequence 51, Appl
26	561	29.0	349	4	US-09-011-769A-47	Sequence 47, Appl
27	560	29.0	613	3	US-09-171-945-113	Sequence 113, App
28	554	28.7	349	4	US-09-011-769A-60	Sequence 60, Appl
29	553	28.6	349	4	US-09-011-769A-64	Sequence 64, Appl
30	553	28.6	417	1	US-07-649-591B-8	Sequence 8, Appli
31	553	28.6	417	1	US-08-277-540-8	Sequence 8, Appli
32	553	28.6	417	1	US-08-430-787A-8	Sequence 8, Appli
33	546	28.3	399	4	US-09-710-099-8	Sequence 8, Appli
34	528.5	27.4	350	4	US-09-675-305-12	Sequence 12, Appl
35	528.5	27.4	419	1	US-07-649-591B-5	Sequence 5, Appli
36	528.5	27.4	419	1	US-08-277-540-5	Sequence 5, Appli
37	528.5	27.4	419	1	US-08-430-787A-5	Sequence 5, Appli
38	519.5	26.9	436	4	US-09-710-099-6	Sequence 6, Appli
39	513	26.6	417	3	US-08-640-906-4	Sequence 4, Appli
40	513	26.6	417	4	US-09-395-936-4	Sequence 4, Appli
41	512	26.5	417	3	US-08-640-906-18	Sequence 18, Appl
42	512	26.5	417	4	US-09-395-936-18	Sequence 18, Appl
43	487.5	25.2	419	3	US-08-640-906-2	Sequence 2, Appli
44	487.5	25.2	419	4	US-09-395-936-2	Sequence 2, Appli
45	482.5	25.0	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Alignment Scores:

Pred. No.:	9.6e-209	Length:	386
Score:	1852.50	Matches:	354
Percent Similarity:	94.16%	Conservative:	1
Best Local Similarity:	93.90%	Mismatches:	5
Query Match:	95.89%	Indels:	18
DB:	4	Gaps:	1

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-813-133A-2 (1-386)

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy       121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy       181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
             |||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||
Db       161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy       541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy       601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy       661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
             ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy       721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy       781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy       841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy       901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
             ||||||||||||||||||||:::|||  ||||||||||||||||||||||||||||||||
Db       301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy       961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy      1021 ATCAAATATTCGTT---------------------------------------------- 1034
             ||||||||||||
Db       341 IleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuPro 360

Qy      1035 ------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
                   ||||||||||||||||||||||||||||||||||||||||||||
Db       361 GluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSerLys 377

RESULT 2
US-07-649-591B-3
; Sequence 3, Application US/07649591B
;
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-3

Alignment Scores:
Pred. No.:                1.15e-206       Length:          423
Score:                    1835.00         Matches:         356
Percent Similarity:       86.23%          Conservative:    1
Best Local Similarity:    85.99%          Mismatches:      3
Query Match:              94.98%          Indels:          55
DB:                       1               Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-07-649-591B-3 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
```

```
Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-3

Alignment Scores:
Pred. No.:                1.15e-206       Length:          423
Score:                    1835.00         Matches:         356
Percent Similarity:       86.23%          Conservative:    1
Best Local Similarity:    85.99%          Mismatches:      3
Query Match:              94.98%          Indels:          55
DB:                       1               Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-277-540-3 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
```

```
Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 4
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-3

Alignment Scores:
Pred. No.:                1.15e-206       Length:          423
Score:                    1835.00         Matches:         356
Percent Similarity:       86.23%          Conservative:    1
Best Local Similarity:    85.99%          Mismatches:      3
Query Match:              94.98%          Indels:          55
DB:                       1               Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-430-787A-3 (1-423)

Qy           1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy          61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy         121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy         181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy         241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy         301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy         361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy         421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT-------------------------------------------------------TAC 1037
              |||                                                       |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  Plasma
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  23..401
US-08-869-057-2

Alignment Scores:
Pred. No.:              1.15e-206      Length:          423
Score:                  1835.00        Matches:         356
Percent Similarity:     86.23%         Conservative:    1
Best Local Similarity:  85.99%         Mismatches:      3
Query Match:            94.98%         Indels:          55
DB:                     2              Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-869-057-2 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
```

```
Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT-------------------------------------------------TAC 1037
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 6
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001173
;  CURRENT APPLICATION NUMBER: US/09/813,133A
;  CURRENT FILING DATE:  2001-06-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Human
US-09-813-133A-4

Alignment Scores:
Pred. No.:              2.25e-205       Length:         423
Score:                  1824.00         Matches:        354
Percent Similarity:     85.75%          Conservative:   1

Best Local Similarity:  85.51%          Mismatches:     5
Query Match:            94.41%          Indels:         55
DB:                     4               Gaps:           2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-813-133A-4 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340
```

```
Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::|||   ||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                      |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              ||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
```

RESULT 7
US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hershberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-139-2

Alignment Scores:

Pred. No.:	2.52e-66	Length:	404
Score:	640.50	Matches:	136
Percent Similarity:	54.29%	Conservative:	60
Best Local Similarity:	37.67%	Mismatches:	122
Query Match:	33.15%	Indels:	43
DB:	1	Gaps:	6

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-696-139-2 (1-404)

```
Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||:::     :::|||                                 :::   :::|||
Db         10 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerLeuLeu 29

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              |||   :::|||   :::|||     |||:::|||   :::         |||
Db         30 HisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGlnIle 49

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                  |||   |||   |||   |||   |||:::     |||:::
Db         50 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 69

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
              |||       :::   :::         ||||||:::   ::::::      ::::::
Db         70 PheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerValLeu 89

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
              :::   |||     :::         |||   :::         ||||||:::
Db         90 GluAlaGlnPheAspSerArgVal---------ArgThrThrGlyHisSerTyrGluLys 106

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
              |||::::::         |||   :::|||   :::   :::|||       :::||||||
Db        107 TyrAsnAsnTrpGluThrIleGluAlaTrpThrLysGlnValThrSerGluAsnProAsp 126

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
              ::::::::::       ||||||:::::|||           :::|||:::||||||
Db        127 LeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeuLys 146

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
              |||   ||||||         |||   ||||||::::::|||||||||   ||||||
Db        147 Val---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHisAla 165

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------- 585
              |||||||||||||||   |||||||||   ||||||:::
Db        166 ArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThrTyr 185

Qy        585 ---------- 585

Db        186 GlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuProVal 205

Qy        586 ----------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGT 615
                      |||:::                 ||||||||||||||||||   |||
Db        206 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 225

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||       |||         ||||||||||||||   |||   ||||||
Db        226 SerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---AspAla 244

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
              ||||||      |||||||||:::       |||   ||||||||||||||
Db        245 GlyTrpCysThrThrGlyAlaSerThrAspProCysAspGluThrTyrCysGlySerAla 264

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
              |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        265 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSerSer 284

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              ||||||||||||:::::::::|||||||||||||||   |||:::::::||||||||||||
Db        285 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 304

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
                        :::::::::   ||||||:::   :::|||       ||||||:::
Db        305 AspTyrLysLeuProGluAsnAsnAlaGluLeuAsnAsnLeuAlaLysAlaAlaValLys 324

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
              :::   |||         |||:::||||||:::|||   |||:::   |||:::
Db        325 GluLeu---AlaThrLeuTyrGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 343

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||   |||   ||||||   |||||||||   ||||||   ||||||||||||||||||
Db        344 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrSerPhe 363
```

```
Qy       1036 ACA 1038
              |||
Db        364 Thr 364

RESULT 8
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  415 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-860-882A-57

Alignment Scores:
Pred. No.:              3.35e-66        Length:         415
Score:                  639.50          Matches:        140
Percent Similarity:     51.72%          Conservative:   71
Best Local Similarity:  34.31%          Mismatches:     136
Query Match:            33.10%          Indels:         62
DB:                     2               Gaps:           8

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-860-882A-57 (1-415)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 72
              |||   |||||||||   :::   |||           |||           |||:::
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         73 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 132
                 .:::|||                              :::   :::::::::
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        133 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 192
                                                                  |||   :::|||   :::|||      |||:::|||   :::          |||
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        193 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 252
                  |||   |||   |||   |||   |||          |||:::      |||
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        253 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 312
                 :::   :::          |||||:::::::::::   ::::::::::::
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        313 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
                 |||      :::               ||||||:::        ||||||:::|||:::
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                      |||   :::|||   :::   :::          :::|||   ::::::
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
                 ::::::      ||||||::::::||||||          :::|||:::|||||||||
Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
                 ||||||           |||   ||||||::::::|||||||||   ||||||||||||
Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 585
                 ||||||||||||||||||||   ||||||:::
Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        585 ------------------------------------------------------------ 585

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        586 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 621
                    |||:::                 :::|||   |||||||||   ||||||
Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                 :::      :::   ||||||||||||||   |||   ||||||           |||
Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
                 ||||||   |||||||||   :::   |||   ||||||||||||||           |||
Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
                 ||||||   |||   ||||||:::|||   |||:::|||      ::::::   ||||||
Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
                 ||||||:::::::::|||||||||||||||   :::::::::||||||||||||
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                      :::::::::   ||||||:::   :::|||          |||:::   :::
Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                    :::          |||:::||||||:::|||   |||:::   |||:::|||
Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1040
                 |||   ||||||   |||||||||   ||||||   ||||||:::||||||||||||
Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1041 ----------------------------------------------------AAACCCACC 1049
                                                                    :::   |||
Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
                 |||   |||   |||   |||:::
```

```
Db        396 CysGluGluThrPheLeuAlaIle 403

RESULT 9
US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 39:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 415 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

Alignment Scores:
Pred. No.:              3.35e-66        Length:         415
Score:                  639.50          Matches:        140
Percent Similarity:     51.72%          Conservative:   71
Best Local Similarity:  34.31%          Mismatches:     136
Query Match:            33.10%          Indels:         62
DB:                     4               Gaps:           8

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-011-769A-39 (1-415)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG---TTCCAG 72
              |||   |||||||||   :::   |||           |||           |||:::
Db          2 LeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGluHisPheGlu 21

Qy         73 AGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAAT 132
                 :::|||                          :::   :::::::::
Db         22 GlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIleArgGlu 41

Qy        133 CTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAG 192
              |||   :::|||   :::|||      |||:::|||   :::         |||
Db         42 LeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLysPro 61

Qy        193 AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTA 252
                 |||   |||   |||   |||   |||           |||:::      |||
Db         62 HisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsnValLeu 81

Qy        253 AATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAG 312
                 :::   :::         ||||||:::::::::::   ::::::::::::
Db         82 LysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValValGluAla 101

Qy        313 CAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
              |||      :::               ||||||:::         ||||||:::|||:::
Db        102 GlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLysTyrAsn 118

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                       |||   :::|||   :::   :::           :::|||   ::::::
Db        119 LysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAlaLeuIle 138

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
              ::::::      ||||||:::::|||||          :::|||:::|||||||||
Db        139 SerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLysVal--- 157

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
              ||||||            |||   ||||||:::::|||||||||   ||||||||||
Db        158 GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArgGlu 177

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 585
              ||||||||||||||||||||   ||||||:::
Db        178 TrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyrGlyArg 197

Qy        585 ---------------------------------------------------------- 585

Db        198 GluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValLeuAsn 217

Qy        586 ------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCTTTC 621
                    |||:::                 :::|||   |||||||||   ||||||
Db        218 IleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArgSerThr 237

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
              :::       :::   ||||||||||||||   |||   ||||||          |||
Db        238 HisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAlaGlyTrp 256

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
              ||||||   |||||||||   :::   |||   ||||||||||||||          |||
Db        257 CysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAlaGlu 276

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
              ||||||   |||   ||||||:::|||   |||:::|||       ::::::   ||||||
Db        277 SerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSerIleLys 296

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
              ||||||:::::::::|||||||||||||||   :::::::::|||||||||||
Db        297 AlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyrAlaTyr 316

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                    :::::::::   ||||||:::   :::|||         |||:::   :::
Db        317 LysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGluLeu 336

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                  :::         |||:::||||||:::|||   |||:::   |||:::|||
Db        337 ---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyrPro 355

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATC- 1040
              |||   ||||||   |||||||||   ||||||   ||||||:::||||||||||||
Db        356 AlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPheThrPhe 375

Qy       1041 ------------------------------------------------AAACCCACC 1049
                                                                :::   |||
Db        376 GluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArgAlaThr 395

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
              |||   |||   |||   |||:::
Db        396 CysGluGluThrPheLeuAlaIle 403

RESULT 10
```

```
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-7

Alignment Scores:
Pred. No.:                5.04e-66      Length:          417
Score:                    638.00        Matches:         147
Percent Similarity:       51.23%        Conservative:    62
Best Local Similarity:    36.03%        Mismatches:      135
Query Match:              33.02%        Indels:          65
DB:                       1             Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-07-649-591B-7 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 75
              :::||||||:::         :::               |||          |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         76 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 135
                 :::|||          :::      :::          ||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        136 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 195
               |||   :::   |||:::      |||   |||                |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        196 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 255
                  |||   |||   |||:::      :::          :::::::::   |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        256 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 315
                    |||          :::|||:::   |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        316 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
                       |||   ::::::   |||   |||          |||   :::|||:::

Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                  ::::::|||   ||||||   |||   :::   |||:::||||||:::|||:::
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
                  ::::::|||   |||||||||:::   |||      |||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
                  ||||||:::         :::   |||||::::::|||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 591
                  ||||||||||||||||||||   ||||||:::                   |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        592 AAT---------------------------------------------------------- 594
                  |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        595 ------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 621
                                           ||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                  |||:::   ||||||||||||||||||||   ||||||          |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
                            ::::::   |||          |||   |||      ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
                  ||||||   |||   |||||||||   :::|||:::|||   ::::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
                  |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                        :::|||::::::|||      ||||||      |||      |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                  |||          ||||||||   :::|||      :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1037
                  |||          ||||||:::|||||||||||||||||:::::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1038 ----------------------------------------------ATCAAACCCACC 1049
                                                            ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
                  |||:::|||          :::|||
Db        398 CysLysGluThrMetLeuSerVal 405

RESULT 11
US-08-277-540-7
; Sequence 7, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
```

```
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-7

Alignment Scores:
Pred. No.:                5.04e-66      Length:          417
Score:                    638.00        Matches:         147
Percent Similarity:       51.23%        Conservative:    62
Best Local Similarity:    36.03%        Mismatches:      135
Query Match:              33.02%        Indels:          65
DB:                       1             Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-277-540-7 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 75
              :::||||||:::          :::                |||          |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         76 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 135
                 :::|||          :::      :::         |||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        136 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 195
                 |||   :::   |||:::      |||   |||                |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        196 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 255
                    |||   |||   |||:::      :::          :::::::::   |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        256 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 315
                       |||         :::|||:::   |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        316 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
                      |||   ::::::   |||   |||          |||   :::|||:::
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                    ::::::|||   ||||||   |||   :::   |||:::||||||:::|||:::

Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
              ::::::|||   |||||||||:::   |||       |||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
              ||||||:::         :::   ||||||::::::|||||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 591
              ||||||||||||||||||||   ||||||:::                      |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        592 AAT---------------------------------------------------------- 594
              |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        595 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 621
                                             ||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                  |||:::   ||||||||||||||||||||   ||||||              |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
                                 ::::::   |||      |||   |||   ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
              ||||||   |||   |||||||||   :::|||:::|||   ::::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
              |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                        :::|||::::::|||      ||||||      |||      |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                    |||         |||||||||   :::|||      :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1037
                    |||         |||||||:::|||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1038 ----------------------------------------------ATCAAACCCACC 1049
                                                            ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
              |||:::|||         :::|||
Db        398 CysLysGluThrMetLeuSerVal 405

RESULT 12
US-08-430-787A-7
; Sequence 7, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-7

Alignment Scores:
Pred. No.:                5.04e-66        Length:         417
Score:                    638.00          Matches:        147
Percent Similarity:       51.23%          Conservative:   62
Best Local Similarity:    36.03%          Mismatches:     135
Query Match:              33.02%          Indels:         65
DB:                       1               Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-430-787A-7 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 75
              :::||||||:::         :::                 |||         |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         76 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 135
                 :::|||          :::      :::          ||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        136 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 195
               |||   :::   |||:::      |||   |||                |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        196 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 255
                  |||   |||   |||:::      :::          :::::::::   |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        256 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 315
                    |||         :::|||:::   |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        316 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
                     |||   ::::::   |||   |||          |||   :::|||:::
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                  :::::::|||   ||||||   |||   :::   |||:::||||||:::|||:::
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
                  ::::::|||   |||||||||:::   |||       ||||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
                  ||||||:::          :::   ||||||::::::||||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 591
                  ||||||||||||||||||   ||||||:::                        |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        592 AAT---------------------------------------------------------- 594
                  |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        595 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 621
                                               ||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                    |||:::   ||||||||||||||||||||   ||||||           |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
                              :::::      |||          |||   |||      ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
                  ||||||   |||   |||||||||   :::|||:::|||   ::::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
                  |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                        :::|||:::::::|||      ||||||      |||      |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                  |||       |||||||||   :::|||         :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1037
                  |||         ||||||:::|||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1038 ----------------------------------------------ATCAAACCCACC 1049
                                                            ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
                  |||:::|||         :::|||
Db        398 CysLysGluThrMetLeuSerVal 405

RESULT 13
US-07-649-591B-6
; Sequence 6, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
```

```
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-6

Alignment Scores:
Pred. No.:                1.14e-65      Length:          417
Score:                    635.00        Matches:         139
Percent Similarity:       52.31%        Conservative:    76
Best Local Similarity:    33.82%        Mismatches:      132
Query Match:              32.87%        Indels:          65
DB:                       1             Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-07-649-591B-6 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
              :::   :::::::::|||:::   |||                       |||
Db          1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||        :::|||        |||:::      :::|||      ::::::
Db         21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              :::::::|||      |||   |||:::      |||   |||                :::
Db         41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                          |||   |||   |||:::      :::         :::::::::
Db         61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                 |||:::   :::   :::          :::|||:::   |||:::::::::   |||
Db         81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy        307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              :::::::|||            :::   :::      |||   |||         |||
Db        101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db        118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::|||:::::::::|||   ||||||||:::   |||      ||||||||||||||
Db        138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::   |||::::::         :::   ||||||::::::||||||||||||||||
Db        158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
              ||||||||||||:::|||||||||||||||   ||||||:::
```

```
Db        177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 ||||||||||||||||||||
Db        217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||         |||:::   ||||||||||||||||||||   ||||||   :::
Db        237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                    |||               ::::::      |||::::::   |||   |||
Db        256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                 ||||||||||   |||   ||||||||   :::|||:::|||   ::::::|||
Db        276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||                 :::||||||:::|||:::   ||||||
Db        316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
                    :::      |||        |||||||||   :::|||                |||
Db        336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              :::|||         |||         ||||||   ||||||||||||||||||:::::
Db        355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy       1033 TTTAC----------------------------------------------------ATC 1040
                |||                                                    |||
Db        375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              ||||||||||||||||||         ||||||
Db        395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 14
US-08-277-540-6
; Sequence 6, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-6

Alignment Scores:
Pred. No.:                1.14e-65        Length:         417
Score:                    635.00          Matches:        139
Percent Similarity:       52.31%          Conservative:   76
Best Local Similarity:    33.82%          Mismatches:     132
Query Match:              32.87%          Indels:         65
DB:                       1               Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-277-540-6 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
              :::   :::::::::|||:::  |||                 |||
Db          1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         :::|||           |||:::     :::|||     ::::::
Db         21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              ::::::|||     |||   |||:::     |||  |||                :::
Db         41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                       |||   |||   |||:::     :::          :::::::::
Db         61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                ||:::   :::   :::        :::|||:::   |||:::::::::   |||
Db         81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy        307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::|||           :::   :::      |||   |||          |||
Db        101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::     :::|||   :::|||   |||   :::   :::::::::|||
Db        118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::|||:::::::::|||   |||||||||:::   |||      ||||||||||||||
Db        138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::   |||::::::          :::   ||||||::::::|||||||||||||||
Db        158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy        585 ---------------------------------------------------------- 585

Db        197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                             |||:::                 ||||||||||||||||||||
Db        217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||         |||:::  ||||||||||||||||||||   ||||||   :::
Db        237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                 |||          ::::::     |||::::::   |||   |||
Db        256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                 ||||||||||||   |||   |||||||||   :::|||:::|||   ::::::|||
Db        276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||          :::||||||:::|||:::   ||||||
Db        316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
                  :::      |||         |||||||||   :::|||           |||
Db        336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              :::|||          |||         ||||||   |||||||||||||||||:::::
Db        355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy       1033 TTTAC-----------------------------------------------------ATC 1040
              |||                                                     |||
Db        375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              ||||||||||||||||||        ||||||
Db        395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 15
US-08-430-787A-6
; Sequence 6, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
```

; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 417 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-6

Alignment Scores:

Pred. No.:	1.14e-65	Length:	417
Score:	635.00	Matches:	139
Percent Similarity:	52.31%	Conservative:	76
Best Local Similarity:	33.82%	Mismatches:	132
Query Match:	32.87%	Indels:	65
DB:	1	Gaps:	9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-08-430-787A-6 (1-417)

```
Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
              :::   :::::::::|||:::   |||                        |||
Db          1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||        :::|||        |||:::     :::|||     ::::::
Db         21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              ::::::|||      |||   |||:::      |||   |||                :::
Db         41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                          |||   |||   |||:::      :::         :::::::::
Db         61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                ||:::   :::   :::          :::|||:::   |||:::::::   |||
Db         81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy        307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              :::::::|||            :::   :::      |||   |||         |||
Db        101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db        118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::|||:::::::::|||   |||||||||:::   |||      ||||||||||||||
Db        138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::   |||::::::         :::   ||||||::::::|||||||||||||||
Db        158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 ||||||||||||||||||||
Db        217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||        |||:::   |||||||||||||||||||   ||||||   :::
Db        237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                  |||               ::::::      |||::::::   |||   |||
Db        256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                |||||||||||   |||   ||||||||   :::|||:::|||   ::::::|||
Db        276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||                :::||||||:::|||:::   ||||||
Db        316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
                    :::   |||   |||||||||   :::|||                  |||
Db        336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              :::|||         |||         ||||||   ||||||||||||||||||:::::
Db        355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy       1033 TTTAC----------------------------------------------------ATC 1040
              |||                                                      |||
Db        375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              ||||||||||||||||    ||||||
Db        395 LysProThrCysArgGluThrMetLeuAlaVal 405
```

Search completed: January 1, 2004, 19:03:29
Job time : 50.1067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:59:01 ; Search time 47.8326 Seconds
(without alignments)
9060.445 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 1467874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135753_25135/app_query.fasta_1.2958
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980881_@CGN_1_1_148_@runat_31122003_135753_25135
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1835	95.0	423	12	US-10-379-836-17	Sequence 17, Appl
2	1746	90.4	423	12	US-10-379-836-2	Sequence 2, Appli
3	1550.5	80.3	422	12	US-10-379-836-18	Sequence 18, Appl
4	1522.5	78.8	422	12	US-10-379-836-16	Sequence 16, Appl
5	839	43.4	211	9	US-09-925-302-467	Sequence 467, App
6	635	32.9	417	12	US-10-341-434-188	Sequence 188, App
7	635	32.9	417	15	US-10-229-546-2	Sequence 2, Appli
8	635	32.9	417	15	US-10-229-546-9	Sequence 9, Appli
9	632	32.7	402	12	US-10-379-836-20	Sequence 20, Appl
10	623	32.2	716	10	US-09-910-059-125	Sequence 125, App
11	613.5	31.8	437	12	US-10-274-639-12	Sequence 12, Appl
12	613.5	31.8	437	14	US-10-200-344-10	Sequence 10, Appl
13	564.5	29.2	374	9	US-09-888-615-61	Sequence 61, Appl
14	563	29.1	444	16	US-10-176-306-74	Sequence 74, Appl
15	560	29.0	613	10	US-09-910-059-113	Sequence 113, App
16	546	28.3	399	15	US-10-200-910-8	Sequence 8, Appli
17	528.5	27.4	350	14	US-10-200-344-12	Sequence 12, Appl
18	519.5	26.9	436	15	US-10-200-910-6	Sequence 6, Appli
19	511.5	26.5	436	14	US-10-076-535-2	Sequence 2, Appli
20	508	26.3	417	9	US-09-923-779-150	Sequence 150, App
21	504	26.1	428	9	US-09-925-297-528	Sequence 528, App
22	495.5	25.6	298	12	US-10-379-836-19	Sequence 19, Appl
23	489.5	25.3	406	9	US-09-925-297-517	Sequence 517, App
24	489.5	25.3	419	12	US-10-345-680-65	Sequence 65, Appl
25	483	25.0	421	11	US-09-946-374-234	Sequence 234, App
26	483	25.0	421	12	US-10-015-387A-234	Sequence 234, App
27	483	25.0	421	12	US-10-006-130A-234	Sequence 234, App
28	483	25.0	421	12	US-10-199-672-308	Sequence 308, App
29	483	25.0	421	12	US-10-006-172A-234	Sequence 234, App
30	483	25.0	421	12	US-10-187-749-308	Sequence 308, App
31	483	25.0	421	12	US-10-194-457-308	Sequence 308, App
32	483	25.0	421	12	US-10-184-642-308	Sequence 308, App
33	483	25.0	421	12	US-10-196-747-308	Sequence 308, App
34	483	25.0	421	12	US-10-015-392A-234	Sequence 234, App
35	483	25.0	421	12	US-10-017-253A-234	Sequence 234, App
36	483	25.0	421	12	US-10-173-689-308	Sequence 308, App
37	483	25.0	421	12	US-10-173-690-308	Sequence 308, App
38	483	25.0	421	12	US-10-173-691-308	Sequence 308, App
39	483	25.0	421	12	US-10-173-692-308	Sequence 308, App
40	483	25.0	421	12	US-10-173-694-308	Sequence 308, App
41	483	25.0	421	12	US-10-173-698-308	Sequence 308, App
42	483	25.0	421	12	US-10-173-699-308	Sequence 308, App
43	483	25.0	421	12	US-10-173-707-308	Sequence 308, App
44	483	25.0	421	12	US-10-174-569-308	Sequence 308, App
45	483	25.0	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Alignment Scores:
Pred. No.: 1.19e-176 Length: 423
Score: 1835.00 Matches: 356

```
Percent Similarity:     86.23%          Conservative:  1
Best Local Similarity:  85.99%          Mismatches:    3
Query Match:            94.98%          Indels:        55
DB:                     12              Gaps:          2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-17 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                    |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Papio hamadryas
US-10-379-836-2

Alignment Scores:
Pred. No.:              1.2e-167        Length:         423
Score:                  1746.00         Matches:        338
Percent Similarity:     83.82%          Conservative:   9
Best Local Similarity:  81.64%          Mismatches:     13
Query Match:            90.37%          Indels:         55
DB:                     12              Gaps:           2

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-2 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||  ||||||||||||||||||||||||||||:::||||||||||||||||||||
Db         61 LeuIleGluLysLysLysGlnValHisPhePheValAsnSerSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
```

```
Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||   ||||||||::::::
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluLeuIleThrGluLysTyr 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||:::||||||:::||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerTyrGluLysHisProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaMetTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGlu 200

Qy        591 ------------------------------------------------------------ 591

Db        201 TyrTyrGlyIleIleGlyGluTyrThrAsnLeuLeuArgHisValAspPheTyrValMet 220

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||   ||||||||||||||||||||   ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnArgCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||   ||||||||||:::|||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluAlaLysAlaValAlaAsnPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||   ||||||||||||||   ||||||||||||||   |||||||||
Db        341 ValArgAlaIleGlnLysThrSerLysAsnIleArgTyrThrHisGlyArgGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyAlaAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT-------------------------------------------------------TAC 1037
              |||                                                         |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyLysTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||::::::||||||||||||||||||||||
Db        401 IleLysProThrCysLysAspAlaPheAlaAlaValSerLys 414

RESULT 3
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-379-836-18

Alignment Scores:
Pred. No.:              7.23e-148       Length:         422
Score:                  1550.50         Matches:        302
Percent Similarity:     78.99%          Conservative:   25
Best Local Similarity:  72.95%          Mismatches:     32
Query Match:            80.25%          Indels:         56
DB:                     12              Gaps:           3

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-18 (1-422)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||       |||   :::||||||   |||:::|||:::   |||||||||
Db          1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db         20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
                 |||   |||||||||:::|||||||||||||||||||||||||||||||||:::|||
Db         60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
Db         80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||:::||||||      ||||||||||||||||||||||||||||||||||||||
Db        100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db        120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db        160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy        591 ------------------------------------------------------------ 591

Db        200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 279
```

```
Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
Db       280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
Db       300 AspHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db       320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
Db       340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
Db       360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy      1030 TCGTT----------------------------------------------------TAC 1037
Db       380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399

Qy      1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
Db       400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLys 413

RESULT 4
US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-379-836-16

Alignment Scores:
Pred. No.:              4.92e-145      Length:         422
Score:                  1522.50        Matches:        297
Percent Similarity:     78.02%         Conservative:   26
Best Local Similarity:  71.74%         Mismatches:     36
Query Match:            78.80%         Indels:         56
DB:                     12             Gaps:           3

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-16 (1-422)

Qy         1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
Db         1 MetLysLeuTyrGlyLeuGlyValLeuValAlaIleLeuLeuTyr---GluLysHisGly 19

Qy        61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
Db        20 LeuAlaPheGlnSerGlyHisValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy       121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
Db        40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy       181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
Db        60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAsnServVal 79

Qy       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
Db        80 LysAlaTyrLeuAsnAlaSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
Db       100 LeuIleGlnGlnGlnThrSerAsnAspThrValSerProArgAlaSerSerSerTyrTyr 119

Qy       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
Db       120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
Db       140 ProAspMetLeuGlnLysIleTyrIleGlySerSerTyrGluLysTyrProLeuTyrVal 159

Qy       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
Db       160 LeuLysValSerGlyLysGluHisArgValLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy       541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
Db       180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy       591 ------------------------------------------------------------ 591
Db       200 PheHisGlyLysGluAsnThrTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy       592 -----------------------------------------AATCGAATGTGGAGAAAG 609
Db       220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy       610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
Db       240 AsnArgSerValHisMetAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy       670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db       260 SerLysHisTrpCysGluLysGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 279

Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
Db       280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
Db       300 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db       320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
Db       340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
Db       360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy      1030 TCGTTTACATC---------------------------------------------- 1040
Db       380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgPhe 399

Qy      1041 ---AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
Db       400 IleLysProThrCysAlaGluAlaLeuAlaAlaValSerLys 413

RESULT 5
US-09-925-302-467
```

```
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
;   LENGTH: 211
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-302-467

Alignment Scores:
Pred. No.:              5.76e-76        Length:         211
Score:                  839.00          Matches:        164
Percent Similarity:     82.59%          Conservative:   2
Best Local Similarity:  81.59%          Mismatches:     9
Query Match:            43.43%          Indels:         27
DB:                     9               Gaps:           3

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-925-302-467 (1-211)

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------------GGC 588
              |||||||||        |||   |||   :::
Db          3 TrpIleSer---MetLeuCysArgTrpLeuMetTrpMetValMetAsnTyrSerTrpLys 21

Qy        589 CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACA 648
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         22 LysAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThr 41

Qy        649 GACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCA 708
              |||||||||   ||||||   ||||||||||||||||||||||||||||||||||||||
Db         42 AspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSer 61

Qy        709 TGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCT 768
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAla 81

Qy        769 AGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCC 828
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         82 SerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSer 101

Qy        829 CAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTG 888
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        102 GlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeu 121

Qy        889 TCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTAT 948
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        122 SerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyr 141

Qy        949 ACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATC 1008
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        142 ThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIle 161

Qy       1009 TATGATTTGGGCATCAAATATTCGTT---------------------------------- 1034
              |||||||||||||||||||||||
Db        162 TyrAspLeuGlyIleLysTyrSerPheThrIleGluLeuArgAspThrGlyThrTyrGly 181

Qy       1035 -----------------TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCT 1076
                              ||||||||||||||||||||||||||||||||||||||||||
Db        182 PheLeuLeuProGluArgTyrIleLysProThrCysArgGluAlaPheAlaAlaValSer 201

Qy       1077 AAA 1079
              |||
Db        202 Lys 202
```

```
RESULT 6
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-341-434-188

Alignment Scores:
Pred. No.:              3.04e-55        Length:         417
Score:                  635.00          Matches:        139
Percent Similarity:     52.31%          Conservative:   76
Best Local Similarity:  33.82%          Mismatches:     132
Query Match:            32.87%          Indels:         65
DB:                     12              Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-341-434-188 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
              :::   :::::::::|||:::   |||                    |||
Db          1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         :::|||           |||:::     :::|||      ::::::
Db         21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              ::::::|||       |||   |||:::      |||   |||                :::
Db         41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                          |||   |||   |||:::      :::        :::::::::
Db         61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                 |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db         81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy        307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::|||             :::   :::       |||   |||         |||
Db        101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db        118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::|||:::::::::|||   |||||||||:::   |||       |||||||||||||||
Db        138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::   |||::::::          :::   ||||||::::::|||||||||||||||
Db        158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196
```

```
Qy       585 ------------------------------------------------------------ 585

Db       197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy       586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 ||||||||||||||||||||
Db       217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy       613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
             ||||||        |||:::   ||||||||||||||||||||   ||||||   :::
Db       237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy       673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                 |||                 ::::::      |||::::::   |||   |||
Db       256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy       733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                ||||||||||||   |||   ||||||||   :::|||:::|||   ::::::|||
Db       276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy       793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
             :::||||||   ||||||:::   ||||||||||||||   ::::::||||||||   
Db       296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy       853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
             ||||||                :::||||||:::|||:::   ||||||
Db       316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy       913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
                 :::     |||         ||||||||   :::|||            |||
Db       336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy       973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
             :::|||         |||         ||||||   |||||||||||||||||:::::: 
Db       355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy      1033 TTTAC-----------------------------------------------------ATC 1040
             |||                                                     |||
Db       375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy      1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
             ||||||||||||||||||        ||||||
Db       395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 7
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-2

Alignment Scores:
Pred. No.:              3.04e-55        Length:         417
Score:                  635.00          Matches:        139
Percent Similarity:     52.31%          Conservative:   76
Best Local Similarity:  33.82%          Mismatches:     132
Query Match:            32.87%          Indels:         65

DB:                     15              Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-229-546-2 (1-417)

Qy        16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
             :::   :::::::::|||:::   |||                  |||
Db         1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy        67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
             |||         :::|||           |||:::      :::|||      ::::::
Db        21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy       127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
             ::::::|||       |||   |||:::      |||   |||                :::
Db        41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy       187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                         |||   |||   |||:::      :::      :::::::::
Db        61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy       247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
               |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db        81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy       307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
             ::::::|||         :::   :::      |||   |||          |||
Db       101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy       364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
             :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db       118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy       424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
             :::|||:::::::::|||   |||||||:::   |||       ||||||||||||||
Db       138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy       484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
             |||:::   |||::::::         :::   ||||||:::::||||||||||||||||
Db       158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy       544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
             ||||||||||||:::|||||||||||||||   ||||||:::
Db       177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy       585 ------------------------------------------------------------ 585

Db       197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy       586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 ||||||||||||||||||||
Db       217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy       613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
             ||||||        |||:::   ||||||||||||||||||||   ||||||   :::
Db       237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy       673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                 |||                 ::::::      |||::::::   |||   |||
Db       256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy       733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                ||||||||||||   |||   ||||||||   :::|||:::|||   ::::::|||
Db       276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy       793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
             :::||||||   ||||||:::   ||||||||||||||   ::::::||||||||   
Db       296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy       853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
             ||||||                :::||||||:::|||:::   ||||||
Db       316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy       913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
```

```
Db        336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
Db        355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy       1033 TTTAC----------------------------------------------------ATC 1040
Db        375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
Db        395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 8
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-9

Alignment Scores:
Pred. No.:              3.04e-55      Length:          417
Score:                  635.00        Matches:         139
Percent Similarity:     52.31%        Conservative:    76
Best Local Similarity:  33.82%        Mismatches:      132
Query Match:            32.87%        Indels:          65
DB:                     15            Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-229-546-9 (1-417)

Qy          16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
Db           1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy          67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db          21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy         127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
Db          41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy         187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
Db          61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy         247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
Db          81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy         307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
Db         101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy         364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
Db         118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy         424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
Db         138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy         484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
Db         158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy         544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
Db         177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy         585 ---------------------------------------------------------- 585

Db         197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy         586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
Db         217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy         613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
Db         237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy         673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
Db         256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy         733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
Db         276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy         793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
Db         296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy         853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
Db         316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy         913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
Db         336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy         973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
Db         355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy        1033 TTTAC----------------------------------------------------ATC 1040
Db         375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy        1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
Db         395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 9
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
```

```
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-379-836-20

Alignment Scores:
Pred. No.:              6.06e-55        Length:         402
Score:                  632.00          Matches:        134
Percent Similarity:     52.31%          Conservative:   70
Best Local Similarity:  34.36%          Mismatches:     126
Query Match:            32.71%          Indels:         61
DB:                     12              Gaps:           7

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-379-836-20 (1-402)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||:::      :::|||                                 :::   ::::::
Db          7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIleIle 26

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db         27 ArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIle 46

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                        |||   |||   |||   |||   |||         |||:::
Db         47 LysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGluAsn 66

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                 |||      :::   :::         ||||||:::::::::::   :::::::::
Db         67 ValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnValVal 86

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
              :::   |||      :::                 ||||||:::         ||||||:::
Db         87 GluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGluLys 103

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
              |||:::          |||   :::|||   :::   :::         :::|||
Db        104 TyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnProAla 123

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
              ::::::::::::      ||||||:::::||||||         :::|||:::||||||
Db        124 LeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeuLys 143

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
              |||   ||||||             |||   ||||||::::::|||||||||   ||||||
Db        144 Val---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------- 585
              ||||||||||||||||||||||||||   ||||||:::
Db        163 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThrTyr 182

Qy        585 ---------- 585

Db        183 GlyArgGluIleGlnValThrGluLeuLeuAsnLysLeuAspPheTyrValLeuProVal 202

Qy        586 ------------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGT 615
                          |||:::                 :::|||   |||||||||   |||
Db        203 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThrArg 222

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||   :::      :::   ||||||||||||||   |||   ||||||
Db        223 SerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---AspAla 241

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
              |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db        242 GlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAla 261

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
              |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db        262 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSerSer 281

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              ||||||||||||:::::::::|||||||||||||||||   :::::::::||||||||||
Db        282 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSerTyr 301

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
                          :::::::::   ||||||:::   :::|||         |||:::
Db        302 AlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLys 321

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
                 :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db        322 GluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIle 340

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||   |||   ||||||   |||||||||   ||||||   ||||||:::|||||||||
Db        341 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSerPhe 360

Qy       1036 ACATC----------------------------------------------------AAA 1043
              |||                                                      :::
Db        361 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleArg 380

Qy       1044 CCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                 ||||||   |||   |||   |||:::
Db        381 AlaThrCysGluGluThrPheLeuAlaIle 390

RESULT 10
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody, a
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-125

Alignment Scores:
Pred. No.:              5.8e-54         Length:         716
Score:                  623.00          Matches:        137
Percent Similarity:     51.34%          Conservative:   74
Best Local Similarity:  33.33%          Mismatches:     139
Query Match:            32.25%          Indels:         62
DB:                     10              Gaps:           8

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-910-059-125 (1-716)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              :::   :::|||   |||||||||   :::   |||           |||
Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         67 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
                 |||:::      :::|||                                 :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40
```

```
Qy       124 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 183
Db        41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy       184 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 243
Db        61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy       244 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 303
Db        81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy       304 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
Db       101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy       364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
Db       118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy       424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
Db       138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy       484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
Db       158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy       544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
Db       177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy       585 ------------------------------------------------------------ 585
Db       197 TyrGlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPro 216

Qy       586 ---------------GGCCAT------------------AATCGAATGTGGAGAAAGAAC 612
Db       217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy       613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
Db       237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy       673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
Db       256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

Qy       733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
Db       276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295

Qy       793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
Db       296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

Qy       853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
Db       316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

Qy       913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
Db       336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy       973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
Db       355 IleTyrProSerAlaGlyThrSerLysAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

Qy      1033 TTTACATC-------------------------------------------------- 1040
Db       375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

Qy      1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
```

```
Db       395 ArgAlaThrCysGluGluThrPheLeuAlaIle 405

RESULT 11
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 12
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

Alignment Scores:
Pred. No.:              4.62e-53        Length:         437
Score:                  613.50          Matches:        142
Percent Similarity:     49.76%          Conservative:   67
Best Local Similarity:  33.81%          Mismatches:     140
Query Match:            31.75%          Indels:         72
DB:                     12              Gaps:           11

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-274-639-12 (1-437)

Qy         1 ATGAAGCTTTGC------AGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAG 54
Db        15 LeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro------------GlyHisSer 30

Qy        55 CATGTCTTC------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 108
Db        31 HisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIleProLysThrGlu 50
```

```
Qy        109 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 168
                 :::       |||:::  ::::::        ::::::  |||||||||||
Db         51 GluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeuTrpGlnPro 70

Qy        169 GTAACAGCTGACCTTATTGTGAAG------AAAAAACAAGTCCATTTTTTTGTAAATGCA 222
                 :::       :::   :::                 ||||||       |||
Db         71 SerSerIleSerTyrValSerGluGlyThrValThrAspValHisIleProGlnAsnGly 90

Qy        223 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 282
              |||             :::   |||   |||       :::   |||         ||||||
Db         91 SerArgAla------LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyrLysValLeu 108

Qy        283 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 342
              :::   |||::::::     :::::::::       |||     |||         |||
Db        109 IleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGlnArgAsnArg 128

Qy        343 GCCTCCGCATCG------TACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 396
                 |||   |||         ||||||   ||||||||||||   ||||||   :::|||
Db        129 ArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIleGlnAsnTrp 148

Qy        397 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 456
              :::      :::   :::   |||       ::::::              ||||||  |||
Db        149 MetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIleGlyArgSer 168

Qy        457 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 516
              :::|||        |||::::::||||||:::   |||:::              |||
Db        169 TyrGluGlyArgSerLeuPheIleLeuLysLeu---GlyArgArgSerArgLeuLysArg 187

Qy        517 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 576
              |||:::||||||||||||||||||||||||||||||||||   |||||||||||
Db        188 AlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAlaPheCysGln 207

Qy        577 TGGTTCATA--------------------------------------------------- 585
              ||||||:::
Db        208 TrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArgLysMetLeu 227

Qy        585 ------------------------------------------------------------ 585

Db        228 AsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHisPheSerTrp 247

Qy        586 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 645
                 ::::::|||   |||||||||   ||||||    :::            |||   |||
Db        248 ThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArgCysArgGly 267

Qy        646 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 705
                 |||   |||   |||:::         |||   ||||||:::||||||||||||
Db        268 ValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAlaSerMetHis 286

Qy        706 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 765
                 |||   :::||||||||||||   :::||||||||||||||||||||||||||||
Db        287 ProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluValLysAlaVal 306

Qy        766 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 825
              |||:::||||||||||::::::            |||:::||||||:::|||   |||:::|||
Db        307 AlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPheHisAlaTyr 326

Qy        826 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 885
              :::|||   :::::::::||||||||||||            :::
Db        327 AlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsnPheArgCys 346

Qy        886 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 945
              :::         |||   :::||||||   |||:::   ::::::               |||
Db        347 ValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyrGlyValArg 365

Qy        946 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1005
              |||   :::|||         |||   |||||||||::::::   |||         ||||||
Db        366 TyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSerMetAspTrp 385

Qy       1006 ATCTATGATTTGGGCATCAAATATTCGTTTAC---------------------------- 1037
                 |||      ||||||   |||:::|||
Db        386 AlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThrGlyTyrPhe 405

Qy       1038 -----------------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
```

```
                    ||||||||||||||   |||         ||||||
Db        406 GlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMetLeuAlaVal 425
```

RESULT 12
US-10-200-344-10
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 437
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-200-344-10

```
Alignment Scores:
Pred. No.:                4.62e-53      Length:          437
Score:                    613.50        Matches:         142
Percent Similarity:       49.76%        Conservative:    67
Best Local Similarity:    33.81%        Mismatches:      140
Query Match:              31.75%        Indels:          72
DB:                       14            Gaps:            11
```

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-200-344-10 (1-437)

```
Qy          1 ATGAAGCTTTGC------AGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAG 54
              :::   ||||||         |||   :::|||   |||
Db         15 LeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro------------GlyHisSer 30

Qy         55 CATGTCTTC------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 108
              |||::::::         :::         :::|||:::      :::|||:::|||
Db         31 HisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIleProLysThrGlu 50

Qy        109 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 168
                 :::       |||:::  ::::::        ::::::  |||||||||||
Db         51 GluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeuTrpGlnPro 70

Qy        169 GTAACAGCTGACCTTATTGTGAAG------AAAAAACAAGTCCATTTTTTTGTAAATGCA 222
                 :::       :::   :::                 ||||||       |||
Db         71 SerSerIleSerTyrValSerGluGlyThrValThrAspValHisIleProGlnAsnGly 90

Qy        223 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 282
              |||             :::   |||   |||       :::   |||         ||||||
Db         91 SerArgAla------LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyrLysValLeu 108

Qy        283 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 342
              :::   |||::::::     :::::::::       |||     |||         |||
Db        109 IleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGlnArgAsnArg 128

Qy        343 GCCTCCGCATCG------TACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 396
                 |||   |||         ||||||   ||||||||||||   ||||||   :::|||
Db        129 ArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIleGlnAsnTrp 148

Qy        397 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 456
              :::      :::   :::   |||       ::::::              ||||||  |||
Db        149 MetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIleGlyArgSer 168

Qy        457 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 516
              :::|||        |||::::::||||||:::   |||:::              |||
```

```
Db        169 TyrGluGlyArgSerLeuPheIleLeuLysLeu---GlyArgArgSerArgLeuLysArg 187

Qy        517 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 576
              |||:::||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        188 AlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAlaPheCysGln 207

Qy        577 TGGTTCATA---------------------------------------------------- 585
              |||||||:::
Db        208 TrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArgLysMetLeu 227

Qy        585 ---------------------------------------------------------- 585

Db        228 AsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHisPheSerTrp 247

Qy        586 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 645
                 ::::::|||   |||||||||   ||||||      :::          |||   |||
Db        248 ThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArgCysArgGly 267

Qy        646 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 705
                 |||   |||   |||:::      |||   ||||||:::||||||||||||
Db        268 ValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAlaSerMetHis 286

Qy        706 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 765
                 |||   :::||||||||||||   :::||||||||||||||||||||||||||||
Db        287 ProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluValLysAlaVal 306

Qy        766 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 825
              |||:::|||||||||::::::          |||:::|||||::|||   |||:::|||
Db        307 AlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPheHisAlaTyr 326

Qy        826 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 885
              :::|||   :::::::::||||||||||||      :::          :::
Db        327 AlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsnPheArgCys 346

Qy        886 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 945
              :::          |||   :::||||||   |||:::   ::::::             |||
Db        347 ValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyrGlyValArg 365

Qy        946 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1005
              |||   :::|||      |||   |||||||||::::::   |||          ||||||
Db        366 TyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSerMetAspTrp 385

Qy       1006 ATCTATGATTTGGGCATCAAATATTCGTTTAC---------------------------- 1037
                 |||       ||||||   |||:::|||
Db        386 AlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThrGlyTyrPhe 405

Qy       1038 ----------------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                                      ||||||||||||||   |||        ||||||
Db        406 GlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMetLeuAlaVal 425

RESULT 13
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Homo sapiens

US-09-888-615-61

Alignment Scores:
Pred. No.:              4.01e-48        Length:         374
Score:                  564.50          Matches:        115
Percent Similarity:     53.70%          Conservative:   52
Best Local Similarity:  36.98%          Mismatches:     87
Query Match:            29.22%          Indels:         58
DB:                     9               Gaps:           6

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-888-615-61 (1-374)

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAA 363
              :::|||:::|||   :::   :::|||||||||   :::      :::|||   |||
Db         29 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 48

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              ||||||   :::   |||||||||   |||:::      |||:::|||::::::
Db         49 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 68

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              ::::::::::|||:::      :::|||   :::::|||   :::|||:::|||   |||
Db         69 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu 88

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::|||      |||      ||||||:::||||||||||||||||
Db         89 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 108

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
              |||||||||||||:::||||||||||||   ||||||:::
Db        109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128

Qy        585 ---------------------------------------------------------- 585

Db        129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 148

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 :::|||:::|||||||||:::
Db        149 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 168

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||   :::   |||      |||   |||||||||||||||   ||||||   :::
Db        169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                  ||||||      |||||||||   :::          :::|||:::||||||
Db        188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                 |||   ||||||||||||   ||||||||||||||||||:::           :::
Db        208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
                 |||      :::::::::||||||||||||   |||   |||:::   ||||||
Db        228 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db        248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
              |||:::   |||          |||   |||      |||      ::::::
Db        268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 286

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              ||||||   :::   |||          ||||||      |||:::||||||   :::|||
Db        287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 306

Qy       1033 TTTACATC---------------------------------------------------- 1040
              :::|||
Db        307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
```

```
            :::||||||||||   |||         ||||||
Db        327 GlnProThrCysGluGluThrMetGluAlaVal 337

RESULT 14
US-10-176-306-74
; Sequence 74, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 444
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-306-74

Alignment Scores:
Pred. No.:              5.97e-48      Length:          444
Score:                  563.00        Matches:         109
Percent Similarity:     56.74%        Conservative:    51
Best Local Similarity:  38.65%        Mismatches:      82
Query Match:            29.14%        Indels:          40
DB:                     16            Gaps:            5

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-10-176-306-74 (1-444)

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAA 363
              :::|||:::|||   :::   :::|||||||||   :::       :::|||   |||
Db        159 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 178

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              ||||||   :::  |||||||||   |||:::       |||:::|||::::::
Db        179 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 198

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::::::::|||:::      :::|||   ::::::|||   :::|||:::|||   |||
Db        199 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProIleTyrTyrLeu 218

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::|||                  |||      ||||||:::|||||||||||||||
Db        219 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 238

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
              |||||||||||||||:::|||||||||||   ||||||:::
Db        239 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 258

Qy        585 ------------------------------------------------------------ 585

Db        259 HisLysAspAsnSerArgIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 278

Qy        586 ---------------GGCCAT------------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                  :::|||:::|||||||||:::
Db        279 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 298

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||   :::   |||      |||   |||||||||||||||   ||||||   :::
Db        299 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 317

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                  ||||||      |||||||||   :::          :::|||:::||||||
Db        318 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 337

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
              |||   ||||||||||||   |||||||||||||||||:::            :::
Db        338 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 357

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              |||      :::::::::||||||||||   |||   |||:::   ||||||
Db        358 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 377

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db        378 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 397

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
              |||:::   |||            |||   |||      |||      ::::::
Db        398 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 416

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              ||||||   :::   |||      ||||||      |||:::||||||   :::|||
Db        417 LeuTyrAlaSerSerGlySerSerArgAspTrpArgAspIleGlyIleProPheSer 436

Qy       1033 TTTACA 1038
              :::|||
Db        437 TyrThr 438

RESULT 15
US-09-910-059-113
; Sequence 113, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody, a
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
```

```
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
;   LENGTH: 613
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised Fd mutant HCPB sequence
US-09-910-059-113

Alignment Scores:
Pred. No.:              1.31e-47        Length:         613
Score:                  560.00          Matches:        134
Percent Similarity:     45.11%          Conservative:   55
Best Local Similarity:  31.98%          Mismatches:     114
Query Match:            28.99%          Indels:         117
DB:                     10              Gaps:           11

US-09-980-881A-1_COPY_18_1097 (1-1080) x US-09-910-059-113 (1-613)

Qy          3 GAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTT 62
              |||       |||          ||||||                  |||
Db        238 GluLeuLysThrProLeuGlyAspThrThrHis------------------ThrCysPro 251

Qy         63 CGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGT 122
              |||   ||||||           ||||||
Db        252 ArgCysProGlu------ProLysSerCys---------------------------- 259

Qy        123 TCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCT 182
                                                        ::::::   |||
Db        260 ------------------------------------------------AspThrProPro 263

Qy        183 TATTGTGAAGAAAAAACAAGTCCATTTTTTTGT--------------------AAATGC 221
                |||          |||      |||                       :::|||
Db        264 ProCysProArgCysProGluProLysSerCysAspThrProProProCysProArgCys 283

Qy        222 ATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTT 281
                  :::   |||::::::|||      |||   |||    |||          ||
Db        284 ProGluProLysSerCysAspThrProProProCysProArg-CysProAlaProGluLe 303

Qy        282 GCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCG 341
              ||||
Db        303 uLeuGlyGly-----------------------------------------------  306

Qy        342 AGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGA 401
               |||:::       ||||||:::|||:::          |||  :::|||   ::
Db        307 -AlaThrGlyHisSerTyrGluLysTyrAsnLysTrpGluThrIleGluAlaTrpThrGl 326

Qy        402 ATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGA 461
              :   :::         :::|||   ::::::::::::      ||||||::::::|||||
Db        326 nGlnValAlaThrGluAsnProAlaLeuIleSerArgSerValIleGlyThrThrPheGl 346

Qy        462 GAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCAT 521
              |         :::|||:::|||||||||   ||||||              |||   |||||
Db        346 uGlyArgAlaIleTyrLeuLeuLysVal---GlyLysAlaGlyGlnAsnLysProAlaIl 365

Qy        522 ATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTT 581
              |::::::|||||||||   ||||||||||||||||||||||||||||||   |||||
Db        365 ePheMetAspCysGlyPheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPh 385

Qy        582 CATA------------------------------------------------------- 585
              |:::
Db        385 eValArgGluAlaValArgThrTyrGlyArgGluIleGlnValThrGluLeuLeuAspLy 405

Qy        586 ------------------------------------GGCCAT----------------- 591
                                                  |||:::
Db        405 sLeuAspPheTyrValLeuProValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLy 425

Qy        592 -AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGA 650
                 :::|||   |||||||||   ||||||   :::      :::   ||||||||||||||
Db        425 sSerArgPheTrpArgLysThrArgSerThrHisThrGlySerSerCysIleGlyThrAs 445

Qy        651 CCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATG 710
              |   |||   ||||||           |||||||||   |||||||||   :::   ||
Db        445 pProAsnArgAsnPhe---AspAlaGlyTrpCysGluIleGlyAlaSerArgAsnProCy 464

Qy        711 CTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAG 770
              |   ||||||||||||||         |||||||||   |||   ||||||:::|||
Db        464 sAspGluThrTyrCysGlyProAlaAlaGluSerGluLysGluThrLysAlaLeuAlaAs 484

Qy        771 TTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCA 830
               |||:::|||      ::::::   ||||||||||||:::::::::||||||||||||||
Db        484 pPheIleArgAsnLysLeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGl 504

Qy        831 GCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTC 890
              |   :::::::::|||||||||||                :::::::::   ||||||::
Db        504 nMetMetIleTyrProTyrSerTyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAs 524

Qy        891 TCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATAC 950
              :   :::|||          |||:::   :::      :::          |||:::|||||
Db        524 nAlaLeuAlaLysAlaThrValLysGluLeu---AlaSerLeuHisGlyThrLysTyrTh 543

Qy        951 ACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTA 1010
              |:::|||   |||:::   |||:::|||   :::   |||          ||||||   ||
Db        543 rTyrGlyProGlyAlaThrThrIleTyrProSerAlaGlyThrSerLysAspTrpAlaTy 563

Qy       1011 TGATTTGGGCATCAAATATTCGTTTACATC---------------------------- 1040
              ||||   ||||||:::|||||||||||||
Db        563 rAspGlnGlyIleArgTyrSerPheThrPheGluLeuArgAspThrGlyArgTyrGlyPh 583

Qy       1041 ----------------------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                                      :::   ||||||   |||   |||   |||:::
Db        583 eLeuLeuProGluSerGlnIleArgAlaThrCysGluGluThrPheLeuAlaIle 601


Search completed: January  1, 2004, 19:25:51
Job time : 68.8326 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:25:26 ; Search time 21.1685 Seconds
(without alignments)
9812.904 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 566616

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135749_24966/app_query.fasta_1.2958
-DB=PIR_76 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_71_@runat_31122003_135749_24966 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1835	95.0	423	2	A41204	carboxypeptidase B
2	692	35.8	415	2	A32129	carboxypeptidase B
3	638	33.0	417	1	A34487	carboxypeptidase A
4	635	32.9	417	2	A43929	carboxypeptidase A
5	622	32.2	416	1	A42332	carboxypeptidase B
6	581.5	30.1	306	1	CPBOB	carboxypeptidase B
7	555.5	28.8	309	2	A38395	mast cell carboxyp
8	553	28.6	417	2	A32128	carboxypeptidase A
9	533.5	27.6	419	1	CPBOA	carboxypeptidase A
10	528.5	27.4	419	1	CPRTA	carboxypeptidase A
11	508	26.3	417	2	A56171	carboxypeptidase A
12	489.5	25.3	419	1	S29127	carboxypeptidase A
13	437.5	22.6	528	2	T33527	hypothetical prote
14	397	20.5	323	2	T20507	hypothetical prote
15	395.5	20.5	667	2	T33526	hypothetical prote
16	391.5	20.3	527	2	T27572	hypothetical prote
17	369.5	19.1	303	1	CPCYB	carboxypeptidase B
18	362.5	18.8	455	2	T33572	hypothetical prote
19	358	18.5	666	2	T24170	hypothetical prote
20	299.5	15.5	430	2	S48976	ECM14 protein - ye
21	277.5	14.4	424	2	S17571	carboxypeptidase T
22	271	14.0	451	2	S20723	carboxypeptidase (
23	251	13.0	497	2	T40260	carboxypeptidase p
24	244.5	12.7	373	2	T26030	hypothetical prote
25	175.5	9.1	999	2	T36021	probable zinc-bind
26	135.5	7.0	1446	2	T30916	carboxypeptidase D
27	130.5	6.8	558	2	S75104	hypothetical prote
28	118.5	6.1	707	2	T40070	origin recognition
29	105.5	5.5	985	2	D82776	pyruvate dehydroge
30	103	5.3	1216	2	H84629	hypothetical prote
31	101.5	5.3	985	2	T29910	hypothetical prote
32	100	5.2	376	1	E69957	gamma-D-glutamyl-L
33	99.5	5.2	1389	2	I50090	carboxypeptidase g
34	99	5.1	1587	2	AB2012	hypothetical prote
35	98.5	5.1	983	2	B49284	immediate-early pr
36	98.5	5.1	1078	2	T44232	hypothetical prote
37	98	5.1	491	2	B96739	hypothetical prote
38	97	5.0	663	2	T37772	telomere length re
39	96.5	5.0	537	2	C64432	hypothetical prote
40	96	5.0	368	2	T21748	hypothetical prote
41	95.5	4.9	355	2	C83850	gamma-D-glutamyl-L
42	95	4.9	444	2	S51199	tryptophan 5-monoo
43	95	4.9	662	2	G86210	hypothetical prote
44	94	4.9	606	2	E90536	hypothetical prote
45	93.5	4.8	986	2	E90596	restriction-modifi

ALIGNMENTS

RESULT 1
A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N;Alternate names: plasma carboxypeptidase B
C;Species: Homo sapiens (man)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C;Accession: A41204
R;Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A;Title: Isolation, molecular cloning, and partial characterization of a novel carboxype
A;Reference number: A41204; MUID:92042093; PMID:1939207
A;Accession: A41204
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <EAT>
A;Cross-references: GB:M75106; NID:g189686; PID:g189687
C;Genetics:
A;Gene: GDB:CPB2
A;Cross-references: GDB:129546; OMIM:212070
A;Map position: 13q14.11-13q14.11
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-114/Domain: activation peptide #status predicted <ACP>
F;115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Alignment Scores:
Pred. No.: 2.82e-152 Length: 423
Score: 1835.00 Matches: 356
Percent Similarity: 86.23% Conservative: 1
Best Local Similarity: 85.99% Mismatches: 3
Query Match: 94.98% Indels: 55
DB: 2 Gaps: 2

US-09-980-881A-1_COPY_18_1097 (1-1080) x A41204 (1-423)

Qy 1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

```
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT-------------------------------------------------------TAC 1037
              |||                                                         |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
```

RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C;Accession: A32129; S17543
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: A32129
A;Molecule type: DNA
A;Residues: 1-415 <CLA>
A;Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R;Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A;Title: Distribution of manganese in rat pancreas and identification of its primary bin
A;Reference number: S17543; MUID:91378950; PMID:1898371
A;Accession: S17543
A;Molecule type: protein
A;Residues: 109-116,'X',118-130 <KOD>
C;Genetics:
A;Gene: CPB
A;Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;354,376/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	2.35e-52	Length:	415
Score:	692.00	Matches:	150
Percent Similarity:	55.09%	Conservative:	61
Best Local Similarity:	39.16%	Mismatches:	126
Query Match:	35.82%	Indels:	46
DB:	2	Gaps:	7

US-09-980-881A-1_COPY_18_1097 (1-1080) x A32129 (1-415)

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||  |||           |||  |||  :::           |||:::|||
Db          1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGluGluHis--- 19

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
                    |||         :::|||                                 |||
Db         20 ------PheAspGlyAsnArgValTyrArgValSerValHisGlyGluAspHisValAsn 37

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ::::::|||  |||      |||  ||||||     |||:::|||  :::|||
Db         38 LeuIleGlnGluLeuAlaAsnThrLysGluIleAspPheTrpLysProAspSerAlaThr 57

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
                :::               |||  |||  |||  |||  ||||||  ::|||
Db         58 GlnValLysProLeuThrThrValAspPheHisValLysAlaGluAspValAlaAspVal 77

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              :::      |||      :::  :::         ||||||:::::::|||  :::
Db         78 GluAsnPheLeuGluGluAsnGluValHisTyrGluValLeuIleSerAsnValArgAsn 97

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ::::::  |||      :::  |||       |||||||||      |||
Db         98 AlaLeuGluSerGlnPheAspSerHisThr---------ArgAlaSerGlyHisSerTyr 114

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              :::|||:::           |||  :::|||||||:::  :::         :::
```

```
Db        115 ThrLysTyrAsnLysTrpGluThrIleGluAlaTrpIleGlnGlnValAlaThrAspAsn 134

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||:::::|||:::      ||||||::::::||||||         :::||||||
Db        135 ProAspLeuValThrGlnSerValIleGlyThrThrPheGluGlyArgAsnMetTyrVal 154

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||:::   ||||||   :::      |||   ||||||:::||||||||||||
Db        155 LeuLysIle---GlyLysThrArgProAsnLysProAlaIlePheIleAspCysGlyPhe 173

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------- 585
              ||||||||||||||||||||||||||||||||||   ||||||:::
Db        174 HisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArg 193

Qy        585 ------------------------------------------------------------ 585

Db        194 ThrTyrAsnGlnGluIleHisMetLysGlnLeuLeuAspGluLeuAspPheTyrValLeu 213

Qy        586 ------------------GGCCAT------------------AATCGAATGTGGAGAAAG 609
                               |||:::                  :::||||||||||||||
Db        214 ProValValAsnIleAspGlyTyrValTyrThrTrpThrLysAspArgMetTrpArgLys 233

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
                ||||||      |||   :::   |||:::|||         |||   ||||||
Db        234 ThrArgSerThrMetAlaGlySerSerCysLeuGlyValArgProAsnArgAsnPhe--- 252

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              :::      |||||||||   |||||||||   |||   ||||||||||||||||||||
Db        253 AsnAlaGlyTrpCysGluValGlyAlaSerArgSerProCysSerGluThrTyrCysGly 272

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
                  ||||||||||||   |||   ||||||:::|||   |||:::|||   |||:::
Db        273 ProAlaProGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeu 292

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              :::   ||||||||||||:::::::::|||||||||||||||||   :::::::::||||||
Db        293 SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrProTyr 312

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||                  :::::::::|||||||||:::   :::         |||
Db        313 SerTyrAspTyrLysLeuProGluAsnTyrGluGluLeuAsnAlaLeuValLysGlyAla 332

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
                :::   :::      |||      |||:::||||||:::|||   |||:::
Db        333 AlaLysGluLeu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThr 351

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |||:::|||   |||   ||||||   |||||||||   ||||||   ||||||||||||
Db        352 ThrIleTyrProAlaAlaGlyGlySerAspAspTrpSerTyrAspGlnGlyIleLysTyr 371

Qy       1030 TCGTTTACA 1038
              |||||||||
Db        372 SerPheThr 374

RESULT 3
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34487
R;Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A;Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A;Reference number: A34487; MUID:90062123; PMID:2584208
A;Accession: A34487
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:              1.23e-47        Length:         417
Score:                  638.00          Matches:        147
Percent Similarity:     51.23%          Conservative:   62
Best Local Similarity:  36.03%          Mismatches:     135
Query Match:            33.02%          Indels:         65
DB:                     1               Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x A34487 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 75
              :::||||||:::        :::             |||         |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         76 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 135
                 :::|||              :::     :::         ||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        136 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 195
              |||   :::   |||:::      |||  |||                 |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        196 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 255
                  |||   |||   |||:::      :::          :::::::::   |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        256 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 315
                       |||          :::|||:::   |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        316 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
                        |||   ::::::   |||   |||         |||   :::|||:::
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                    :::::::|||   ||||||   |||   :::   |||:::||||||:::|||:::
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
              ::::::|||   |||||||||:::   |||         ||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
              ||||||:::          :::   ||||||::::::||||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 591
              ||||||||||||||||||||   ||||||:::                     |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        592 AAT---------------------------------------------------------- 594
              |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        595 ------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 621
                                           ||||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                  |||:::   ||||||||||||||||||||   ||||||           |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
                           ::::::   |||      |||   |||   ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
              ||||||   |||   |||||||||   :::|||:::|||   :::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
              |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318
```

```
Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                  :::|||::::::|||     ||||||     |||     |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                |||        |||||||||   :::|||     :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1037
                |||        ||||||:::|||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1038 ----------------------------------------------------ATCAAACCCACC 1049
                                                                    ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
              |||:::|||         :::|||
Db        398 CysLysGluThrMetLeuSerVal 405
```

RESULT 4
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N;Alternate names: mast cell carboxypeptidase A3
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: A43929; A39246; A45759
R;Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A;Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A
A;Reference number: A43929; MUID:92105393; PMID:1729276
A;Accession: A43929
A;Molecule type: DNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:M73716
A;Experimental source: mast cell
A;Note: the authors translated the codon CGC for residue 231 as Thr
R;Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafin,
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A;Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison
A;Reference number: A39246; MUID:90083291; PMID:2594780
A;Accession: A39246
A;Molecule type: mRNA
A;Residues: 1-417 <RE2>
A;Cross-references: GB:M27717; NID:g179933; PIDN:AAA35652.1; PID:g179934
R;Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.
J. Clin. Invest. 83, 1630-1636, 1989
A;Title: Human mast cell carboxypeptidase. Purification and characterization.
A;Reference number: A45759; MUID:89214692; PMID:2708524
A;Accession: A45759
A;Molecule type: protein
A;Residues: 110-137 <GOL>
C;Genetics:
A;Gene: GDB:CPA3
A;Cross-references: GDB:125231; OMIM:114851
A;Map position: 3q21.3-3q25
A;Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-109/Domain: activation peptide #status predicted <ACT>
F;110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F;176,179,304/Binding site: zinc (His, Glu, His) #status predicted
F;245-268/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	2.26e-47	Length:	417
Score:	635.00	Matches:	139
Percent Similarity:	52.31%	Conservative:	76
Best Local Similarity:	33.82%	Mismatches:	132
Query Match:	32.87%	Indels:	65

```
DB:                    2             Gaps:          9

US-09-980-881A-1_COPY_18_1097 (1-1080) x A43929 (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
              :::   :::::::::|||:::  |||                  |||
Db          1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||        :::|||           |||:::     :::|||     ::::::
Db         21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              ::::::|||       |||   |||:::     |||   |||                :::
Db         41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                          |||   |||   |||:::     :::     :::::::::
Db         61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                 |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db         81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy        307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::|||           :::   :::       |||   |||         |||
Db        101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db        118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::|||:::::::::|||   |||||||||:::   |||       |||||||||||||
Db        138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::   |||::::::          :::   |||||:::::|||||||||||||||
Db        158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 585
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy        585 ---------------------------------------------------------- 585

Db        197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy        586 --------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 ||||||||||||||||||||
Db        217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||        |||:::  ||||||||||||||||||||  ||||||  :::
Db        237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                 |||                :::::     |||::::::  |||  |||
Db        256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                 ||||||||||||  |||  |||||||||  :::|||:::|||  ::::::|||
Db        276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              :::||||||   ||||||:::   |||||||||||||   ::::::|||||||||
Db        296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuProProTyrGly 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||                :::||||||:::|||:::   ||||||
Db        316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
```

```
Db        336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032

Db        355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy       1033 TTTAC----------------------------------------------------ATC 1040

Db        375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073

Db        395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 5
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N;Alternate names: carboxypeptidase B isozyme 1
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: A42332; S08254; S02813; S02812
R;Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992
A;Title: Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Ident
A;Reference number: A42332; MUID:92129345; PMID:1370825
A;Accession: A42332
A;Molecule type: mRNA
A;Residues: 1-416 <YAM>
A;Cross-references: GB:M81057; NID:g809194; PIDN:AAA66973.1; PID:g189625
A;Experimental source: pancreas
A;Note: sequence extracted from NCBI backbone (NCBIN:78573, NCBIP:78574)
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08254
A;Molecule type: protein
A;Residues: 16-40 <MOU>
A;Experimental source: pancreas
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02813
A;Molecule type: protein
A;Residues: 16-36,'X',38-43 <PAS>
A;Experimental source: pancreas
A;Accession: S02812
A;Molecule type: protein
A;Residues: 16-36,'Q',38-43 <PA2>
A;Experimental source: pancreas
C;Genetics:
A;Gene: GDB:CPB1
A;Cross-references: GDB:132160; OMIM:114852
A;Map position: 3q21.3-3q25
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-110/Domain: activation peptide #status predicted <ACP>
F;111-415/Product: carboxypeptidase B, pancreatic #status predicted <MAT>
F;173-186,258-272/Disulfide bonds: #status predicted
F;176,179,303/Binding site: zinc (His, Glu, His) #status predicted
F;355,377/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:                3.09e-46        Length:          416
Score:                    622.00          Matches:         139
Percent Similarity:       51.58%          Conservative:    73
Best Local Similarity:    33.82%          Mismatches:      137
Query Match:              32.19%          Indels:          63
DB:                       1               Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x A42332 (1-416)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66

Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         67 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123

Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 183

Db         41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        184 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 243

Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy        244 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 303

Db         81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy        304 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363

Db        101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423

Db        118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483

Db        138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543

Db        158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585

Db        177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPro 216

Qy        586 --------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612

Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672

Db        237 ArgSerThrHisThrGlySer---SerIleGlyThrAspProAsnArgAsnPhe---Asp 254

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732

Db        255 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 274

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792

Db        275 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 294

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852

Db        295 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 314

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912

Db        315 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 334

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972

Db        335 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 353
```

```
Qy       973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
             :::|||   |||   ||||||   |||||||||   ||||||   ||||||:::||||||
Db       354 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 373

Qy      1033 TTTACATC-------------------------------------------------------- 1040
             ||||||
Db       374 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 393

Qy      1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
             :::   ||||||   |||   |||   |||:::
Db       394 ArgAlaThrCysGluGluThrPheLeuAlaIle 404
```

RESULT 6
CPBOB
carboxypeptidase B (EC 3.4.17.2) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
C;Accession: A93797; A92150; A00912
R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A;Title: Amino-acid sequence of bovine carboxypeptidase B.
A;Reference number: A93797; MUID:75217824; PMID:1057162
A;Accession: A93797
A;Molecule type: protein
A;Residues: 1-306 <TIT>
R;Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations c
A;Reference number: A92150; MUID:74260705; PMID:4833744
A;Accession: A92150
A;Molecule type: protein
A;Residues: 31-93;131-181;263-265;292-306 <SCH>
R;Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A;Reference number: A92839; MUID:76265065; PMID:957425
A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R;Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptid
A;Reference number: A92051; MUID:70007159; PMID:5344132
A;Contents: annotation; active site
R;Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptida
A;Reference number: A92119; MUID:73061487; PMID:4565668
A;Contents: annotation; active site
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-158,149-163/Disulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

Alignment Scores:

Pred. No.:	1.05e-42	Length:	306
Score:	581.50	Matches:	120
Percent Similarity:	56.66%	Conservative:	46
Best Local Similarity:	40.96%	Mismatches:	70
Query Match:	30.10%	Indels:	58
DB:	1	Gaps:	6

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPBOB (1-306)

```
Qy       358 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 417
             ||||||:::|||:::::      |||   :::|||   |||   :::
Db         6 TyrGluLysTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25

Qy       418 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 477
             :::||||||:::::::::::      ||||||:::::|||            :::|||
Db        26 AsnProAspLeuIleSerArgSerAlaIleGlyThrThrPheLeuGlyAsnThrIleTyr 45

Qy       478 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 537
             :::|||||||||   ||||||      :::   |||   |||:::::::::|||||||||
Db        46 LeuLeuLysVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64

Qy       538 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 585
             ||||||||||||||||||||||||||||||||||||||||||||||   ||||||:::
Db        65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84

Qy       585 ------------------------------------------------------------ 585

Db        85 ArgThrTyrGlyArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104

Qy       586 ----------------------GGCCAT-----------------AATCGAATGTGGAGA 606
                                     |||:::                 ||||||||||||||
Db       105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrThrAsnArgMetTrpArg 124

Qy       607 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 666
             |||   ||||||      |||   :::   |||   ||||||||||||||   ||||||
Db       125 LysThrArgSerThrArgAlaGlySerSerCysThrGlyThrAspLeuAsnArgAsnPhe 144

Qy       667 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 726
                 ||||||      |||||||||::::::   ||||||||||||||||||
Db       145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163

Qy       727 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 786
             |||         |||||||||   |||   ||||||||||||   |||:::|||   :::
Db       164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgAsnHis 183

Qy       787 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 846
             ::::::   ||||||||||||:::::::::||||||||||||||   :::::::::|||
Db       184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203

Qy       847 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 906
             |||||||||               |||::::::   ||||||:::   :::|||
Db       204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223

Qy       907 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 966
             ||||||:::   :::       :::       |||   |||:::::::|||   |||:::
Db       224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242

Qy       967 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1026
             |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db       243 ThrThrIleTyrProAlaSerGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLys 262

Qy      1027 TATTCGTTTACATC-------------------------------------------------- 1040
             ||||||||||||
Db       263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282

Qy      1041 ------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                   :::|||||||||   |||         |||:::
Db       283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIle 295
```

RESULT 7
A38395
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C;Accession: A38395; A33118
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A;Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme activ
A;Reference number: A38395; MUID:91105153; PMID:1988052
A;Accession: A38395
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-309 <COL>
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A;Reference number: A33118
A;Accession: A33118
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-309 <CO2>
C;Superfamily: carboxypeptidase

C;Keywords: hydrolase
F;248,270/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	1.97e-40	Length:	309
Score:	555.50	Matches:	118
Percent Similarity:	54.27%	Conservative:	41
Best Local Similarity:	40.27%	Mismatches:	77
Query Match:	28.75%	Indels:	58
DB:	2	Gaps:	6

US-09-980-881A-1_COPY_18_1097 (1-1080) x A38395 (1-309)

```
Qy        358 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 417
Db          8 TyrAlaLysTyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLys 27

Qy        418 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 477
Db         28 HisProGluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyr 47

Qy        478 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 537
Db         48 ValLeuLysIle---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66

Qy        538 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 585
Db         67 IleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86

Qy        586 ---------GGCCATAAT------------------------------------------ 594
Db         87 LysSerTyrGlyLysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106

Qy        595 ----------------------------------------------CGAATGTGGAGA 606
Db        107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126

Qy        607 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 666
Db        127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146

Qy        667 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 726
Db        147 ---AspValSerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArg 165

Qy        727 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 786
Db        166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185

Qy        787 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 846
Db        186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205

Qy        847 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 906
Db        206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIle 225

Qy        907 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 966
Db        226 AlaThrAspValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAla 244

Qy        967 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1026
Db        245 SerThrIleTyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLys 264

Qy       1027 TATTCGTTTAC---------------------------------------------- 1037
Db        265 HisThrPheAlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSer 284

Qy       1038 ---ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
Db        285 ArgIleLysProThrCysLysGluThrMetLeuSerVal 297
```

RESULT 8
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C;Accession: A32128
R;Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rutt
J. Biol. Chem. 263, 17828-17836, 1988
A;Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evo
A;Reference number: A32128; MUID:89034323; PMID:3182871
A;Accession: A32128
A;Molecule type: mRNA
A;Residues: 1-417 <GAR>
A;Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	3.32e-40	Length:	417
Score:	553.00	Matches:	137
Percent Similarity:	50.37%	Conservative:	69
Best Local Similarity:	33.50%	Mismatches:	137
Query Match:	28.62%	Indels:	66
DB:	2	Gaps:	13

US-09-980-881A-1_COPY_18_1097 (1-1080) x A32128 (1-417)

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
Db          1 MetArgLeuThrLeuLeuLeuAlaAlaLeuLeuGlyTyrIleTyrCysGlnGlu------ 18

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
Db         19 ---ThrPheValGlyAspGlnValLeuGluIleIleProSerHisGluGluGlnIleArg 37

Qy        121 GTTCTACAGAATCTTACTACAACA------TATGAGATTGTTCTCTGGCAGCCGGTAACA 174
Db         38 ThrLeuLeuGlnLeuGluAlaGluGluHisLeuGluLeuAspPheTrpLysSerProThr 57

Qy        175 GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGAC 234
Db         58 ------------IleProGlyGluThrValHisValArgValProPheAlaSerIleGln 73

Qy        235 AATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTG 294
Db         74 AlaValLysValPheLeuGluSerGlnGlyIleAspTyrSerIleMetIleGluAspVal 93

Qy        295 GAAGATCTTATTCAACAGCAGATTTCCAAC---------GACACAGTCAGCCCCCGAGCC 345
Db         94 GlnValLeuLeuAspGlnGluArgGluGluMetLeuPheAsnGlnGlnArgGluArgGly 113

Qy        346 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 405
Db        114 GlyAsnPheAsnPheGluAlaTyrHisThrLeuGluGluIleTyrGlnGluMetAspAsn 133

Qy        406 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 465
Db        134 LeuValAlaGluAsnProGlyLeuValSerLysValAsnLeuGlySerSerPheGluAsn 153

Qy        466 TACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACAGCCAAAAATGCCATA 522
Db        154 ArgProMetAsnValLeuLysPheSerThrGlyGlyAsp---------LysProAlaIle 170

Qy        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGG--- 579
Db        171 TrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAlaThrAlaLeuTrpThr 190

Qy        579 ------------------------------------------------------------ 579
Db        191 AlaAsnLysIleAlaSerAspTyrGlyThrAspProAlaIleThrSerLeuLeuAsnThr 210

Qy        580 ---------TTCATA--------------------GGCCAT----------------- 591
```

```
                    |||:::                          |||:::
Db        211 LeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrValPheSerGlnThrThr 230

Qy        592 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 651
              |||||||||||||||||   ||||||      :::   :::   |||:::|||   |||
Db        231 AsnArgMetTrpArgLysThrArgSerLysArgSerGlySerGlyCysValGlyValAsp 250

Qy        652 CTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 711
              |||   |||:::       ::::::          |||||||||||||||   |||
Db        251 ProAsnArgAsnTrp---AspAlaAsnPheGlyGlyProGlyAlaSerSerSerProCys 269

Qy        712 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 771
              |||::::::|||   |||      |||   ||||||   |||||||||::::::
Db        270 SerAspSerTyrHisGlyProLysProAsnSerGluValGluValLysSerIleValAsp 289

Qy        772 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 831
              |||::::::   :::      ::::::||||||:::|||::::::|||||||||||||||
Db        290 PheIleLysSerHis---GlyLysValLysAlaPheIleThrLeuHisSerTyrSerGln 308

Qy        832 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 891
              ::::::||||||||   |||      :::|||      |||      ||||||
Db        309 LeuLeuMetPheProTyrGlyTyrLysCysThrLysProAspAspPheAsnGluLeuAsp 328

Qy        892 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACA 951
              ||||||   :::|||   :::|||:::   |||            |||   |||
Db        329 GluValAlaGlnLysAlaAlaGlnAlaLeu---LysArgLeuHisGlyThrSerTyrLys 347

Qy        952 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1011
              |||           :::|||   |||   ||||||      ||||||   |||
Db        348 ValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySerIleAspTrpAlaTyr 367

Qy       1012 GATTTGGGCATCAAATATTCGTTTACATCAAAC-------------------------- 1044
              ||||||||||||||||||||||||
Db        368 AspLeuGlyIleLysTyrSerPheAlaPheGluLeuArgAspThrAlaPheTyrGlyPhe 387

Qy       1045 ---CCACCTGTAGAGAAGCTTTTGCCG 1068
                  |||   :::::::::||||||
Db        388 LeuLeuProAlaLysGlnIleLeuPro 396
```

RESULT 9
CPBOA
carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine
N;Alternate names: procarboxypeptidase A activation peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1992 #text_change 05-May-2000
C;Accession: JN0126; JT0440; A31406; S00059; A38834; A00910; A90355
R;Le Hueerou, I.; Guilloteau, P.; Toullec, R.; Puigserver, A.; Wicker, C.
Biochem. Biophys. Res. Commun. 175, 110-116, 1991
A;Title: Cloning and nucleotide sequence of a bovine pancreatic preprocarboxypeptidase A
A;Reference number: JN0126; MUID:91151335; PMID:1998496
A;Accession: JN0126
A;Molecule type: mRNA
A;Residues: 1-419 <LEH>
A;Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R;Foglizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A;Title: Primary structure of the activation peptide from bovine pancreatic procarboxype
A;Reference number: JT0440; MUID:88209583; PMID:3365436
A;Accession: JT0440
A;Molecule type: protein
A;Residues: 17-110 <FOG>
R;Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A;Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidas
A;Reference number: A31406; MUID:89150306; PMID:3147705
A;Accession: A31406
A;Molecule type: protein
A;Residues: 17-110 <WAD>
R;Chapus, C.; Kerfelec, B.; Foglizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A;Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A;Reference number: S00059; MUID:87275921; PMID:3609014
A;Accession: S00059
A;Molecule type: protein
A;Residues: 17-69 <CHA>
R;Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A;Title: The amino acid sequence of bovine carboxypeptidase A.
A;Reference number: A38834; MUID:70050818; PMID:5260942
A;Accession: A38834
A;Molecule type: protein
A;Residues: 111-137,'Q',139-140,'E',142-417 <BRA>
R;Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A;Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterizatio
A;Reference number: A90360; MUID:71155185; PMID:4927805
A;Accession: A00910
A;Molecule type: protein
A;Residues: 214-411 <BR2>
A;Note: this is the final paper in a series
R;Petra, P.H.; Hermodson, M.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 4023-4025, 1971
A;Title: Characterization of bovine carboxypeptidase A(Allan).
A;Reference number: A90355; MUID:72138789; PMID:5143102
A;Accession: A90355
A;Molecule type: protein
A;Residues: 111-142 <PET>
R;Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A;Title: Identification of the amino acid replacements characterizing the allotropic form
A;Reference number: A90556; MUID:69283620; PMID:5817619
A;Contents: annotation; allelic variant
A;Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CPA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137,'Q',139-14
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/Disulfide bonds: #status experimental
F;358,380/Active site: Tyr, Glu #status experimental

Alignment Scores:

Pred. No.:	1.68e-38	Length:	419
Score:	533.50	Matches:	134
Percent Similarity:	49.87%	Conservative:	61
Best Local Similarity:	34.27%	Mismatches:	133
Query Match:	27.61%	Indels:	63
DB:	1	Gaps:	11

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPBOA (1-419)

```
Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              |||      |||:::||||||:::        |||      |||
Db          4 LeuLeuIleLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp------------ 19

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         |||||||||                              :::||||||   :::
Db         20 PheValGlyHisGlnValLeuArgIleThrAlaAlaAspGluAlaGluValGlnThrVal 39

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGG------------CAGCCG 168
              :::   |||                :::   |||      |||            ||||||
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProGlyGlnPro 59

Qy        169 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 228
                                                  :::      |||
Db         60 ----------------------GlySerProIleAspValArgValProPheProSer 71
```

```
Qy        229 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 288
              :::     ||||||     ||     ||||||         :::::::::
Db         72 LeuGlnAlaValLysValPheLeuGluAlaHisGlyIleArgTyrArgIleMetIleGlu 91

Qy        289 GACGTGGAAGATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCC 339
              ||||||:::   |||:::          |||:::|||:::
Db         92 AspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaSerGlnSerArgAla 111

Qy        340 CGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 396
              |||::::::   ::::::   |||       ||||||:::|||:::|||||||||   :::
Db        112 ArgSerThrAsnThrPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrAspPhe 131

Qy        397 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 456
              ::::::   :::       ||||||   ::::::::::|||:::   ||||||   |||
Db        132 MetAspLeuLeuValAlaGluHisProGlnLeuValSerLysLeuGlnIleGlyArgSer 151

Qy        457 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 516
              :::|||      |||:::||||||||||||   |||           :::   :::
Db        152 TyrGluGlyArgProIleTyrValLeuLysPheSer---ThrGlyGlySerAsnArgPro 170

Qy        517 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 576
              |||||||||||||||   |||||||||:::|||||||||||:::   |||      :::
Db        171 AlaIleTrpIleAspLeuGlyIleHisSerArgGluTrpIleThrGlnAlaThrGlyVal 190

Qy        577 TGGTTCATA---------------------------------------------------- 585
              ||||||
Db        191 TrpPheAlaLysLysPheThrGluAspTyrGlyGlnAspProSerPheThrAlaIleLeu 210

Qy        585 ---------------------------------------------------------- 585

Db        211 AspSerMetAspIlePheLeuGluIleValThrAsnProAspGlyPheAlaPheThrHis 230

Qy        586 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 645
              ||||||:::|||||||||   ||||||      :::::::::   |||:::|||
Db        231 SerGlnAsnArgLeuTrpArgLysThrArgSerValThrSerSerSerLeuCysValGly 250

Qy        646 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 705
              |||   |||   |||:::           :::   :::   ||||||||||||||||
Db        251 ValAspAlaAsnArgAsnTrp---AspAlaGlyPheGlyLysAlaGlyAlaSerSerSer 269

Qy        706 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 765
              ||||||||||||||   |||   |||      ||||||   |||||||||::::::
Db        270 ProCysSerGluThrTyrHisGlyLysTyrAlaAsnSerGluValGluValLysSerIle 289

Qy        766 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 825
              |||::::::   :::   |||      ||||||::::::|||:::|||||||||
Db        290 ValAspPheValLysAspHisGlyAsn---PheLysAlaPheLeuSerIleHisSerTyr 308

Qy        826 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 885
              ||||||   :::::::::|||||||   ||||||                |||      |||
Db        309 SerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrGlnSerIleProAspLysThrGlu 328

Qy        886 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 945
              |||:::   ||||||      ||||||   |||:::   |||:::          |||
Db        329 LeuAsnGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyrGlyThrSer 347

Qy        946 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1005
              |||   :::|||          |||:::|||   |||   ||||||      ||||||
Db        348 TyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSerGlyGlySerIleAspTrp 367

Qy       1006 ATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
              |||:::   ||||||||||||||||||||
Db        368 SerTyrAsnGlnGlyIleLysTyrSerPheThr 378
```

RESULT 10
CPRTA
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.L
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982

A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <QUI>
A;Cross-references: GB:J00713; NID:g203364; PIDN:AAA40893.1; PID:g203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA
A;Residues: 1-260,'F',262,'M',264-346,'K',348-419 <CLA>
C;Genetics:
A;Gene: CPA1
A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	4.61e-38	Length:	419
Score:	528.50	Matches:	135
Percent Similarity:	50.00%	Conservative:	61
Best Local Similarity:	34.44%	Mismatches:	133
Query Match:	27.36%	Indels:	63
DB:	1	Gaps:	11

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPRTA (1-419)

```
Qy          4 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 63
              :::|||      |||::::::|||:::      |||          |||
Db          3 ArgLeuLeuIleLeuSerLeuLeuLeuGluAlaValCysGlyAsnGluAsn--------- 19

Qy         64 GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
              |||         |||||||||                                |||||||||
Db         20 ---PheValGlyHisGlnValLeuArgIleSerAlaAlaAspGluAlaGlnValGlnLys 38

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAGCCGGTAACAGCT 177
              ::::::  |||           :::   |||      |||:::
Db         39 ValLysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg 58

Qy        178 GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT 237
              |||             :::|||   |||                    :::   :::
Db         59 AlaGlyIleProIleAspValArgValProPhe------------ProSerIleGlnSer 74

Qy        238 GTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAA 297
              |||||||||   |||         ||||||         :::::::::   ||||||:::
Db         75 ValLysAlaPheLeuGluTyrHisGlyIleSerTyrGluIleMetIleGluAspValGln 94

Qy        298 GATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCC 348
              |||:::         :::||||||:::|||                    ::::::
Db         95 LeuLeuLeuAspGluGluLysGlnGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr 114

Qy        349 GCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 405
              |||:::   |||       ||||||:::|||:::|||||||||   :::::::::
Db        115 AspSerPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrGluPheMetAspLeu 134

Qy        406 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 465
              :::         ||||||   :::::::::||||||   ||||||::::::||||||
Db        135 LeuValAlaGluHisProGlnLeuValSerLysIleGlnIleGlyAsnThrPheGluGly 154

Qy        466 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 525
              |||::::::|||||||||   |||             |||   :::   |||||||||
Db        155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTrp 173

Qy        526 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 585
```

```
                ||||||   |||||||||:::|||||||||::::::   |||      :::||||||
Db        174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193

Qy        586 GGCCAT------------------------------------------------------ 591

Db        194 LysLysValThrLysAspTyrGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213

Qy        592 ----------------------------------------------------------AAT 594
                                                                         |||
Db        214 AspIlePheLeuGluIleValThrAsnProAspGlyPheAlaTyrThrHisLysThrAsn 233

Qy        595 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 654
              ||||||||||||||   ||||||         :::   |||:::|||   |||
Db        234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuCysValGlyValAspPro 253

Qy        655 AATAGCAACTTTGTCTCCAAACACTGG-----------TGTGAGGAAGGTGCATCCAGT 702
              |||   |||               |||              :::   ||||||||||||
Db        254 AsnArgAsn--------------TrpAspAlaGlyLeuGlyLysAlaGlyAlaSerSer 268

Qy        703 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 762
              :::   ||||||||||||||   |||   :::|||   ||||||   |||||||||:::
Db        269 AsnProCysSerGluThrTyrArgGlyLysPheProAsnSerGluValGluValLysSer 288

Qy        763 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 822
              :::      |||:::      :::   |||  ||||||||:::||||||:::||||||
Db        289 IleValAspPheValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSer 307

Qy        823 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 882
              |||||||||   :::::::::||||||   ||||||         :::   |||
Db        308 TyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrSerGluProAlaProAspGlnAla 327

Qy        883 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 942
              ||||||      :::|||      ||||||   |||:::      :::          |||
Db        328 GluLeuAspGlnLeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThr 346

Qy        943 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1002
                 :::   :::|||        :::|||:::|||   |||   |||         |||
Db        347 GluPheLysTyrGlySerIleIleAspThrIleTyrGlnAlaSerGlySerThrIleAsp 366

Qy       1003 TGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
              |||   |||      ||||||||||||||||||||
Db        367 TrpThrTyrSerGlnGlyIleLysTyrSerPheThr 378

RESULT 11
A56171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C;Accession: A56171; S02809; S71395
R;Catasus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAA74425.1; PID:g790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S',18-36,'N',38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LAE>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Alignment Scores:
Pred. No.:              2.85e-36        Length:         417
Score:                  508.00          Matches:        128
Percent Similarity:     48.48%          Conservative:   63
Best Local Similarity:  32.49%          Mismatches:     137
Query Match:            26.29%          Indels:         66
DB:                     2               Gaps:           11

US-09-980-881A-1_COPY_18_1097 (1-1080) x A56171 (1-417)

Qy         55 CATGTCTTC---------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACC 105
              |||::::::          |||      |||||||||     :::|||
Db         13 HisIleTyrCysLeuGluThrPheValGlyAspGlnValLeuGluIleValProSerAsn 32

Qy        106 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC------ 159
                 |||::::::   |||      |||                 :::   |||
Db         33 GluGluGlnIleLysAsnLeuLeuGlnLeuGluAlaGlnGluHisLeuGlnLeuAspPhe 52

Qy        160 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 219
              |||:::     |||                 :::     |||      |||
Db         53 TrpLysSerProThr------------ThrProGlyGluThrAlaHisValArgValPro 68

Qy        220 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 279
                 :::|||      ||||||      |||             ||||||      |||:::
Db         69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTyrSerIle 88

Qy        280 TTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCC 339
              ::::::   ||||||:::   |||:::   ::::::                  :::
Db         89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108

Qy        340 CGAGCCTCCGCATCG---------TACTATGAACAGTATCACTCACTAAATGAAATCTAT 390
              |||          |||            :::      ||||||:::|||   ||||||
Db        109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyrHisThrLeuGluGluIleSer 128

Qy        391 TCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGA 450
                 ::::::   :::          ||||||   :::::::::|||::::::||||||
Db        129 GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148

Qy        451 TCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACA 507
              ||||||||||||      |||:::   |||||||||   |||   |||   :::
Db        149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp------ 166

Qy        508 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 567
                 |||   |||||||||:::|||   |||||||||||||||||||:::::   |||
Db        167 ---LysProAlaIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185

Qy        568 TTCTGCTTGTGGTTCATAGGC---------------------------------------- 588
                    ||||||
Db        186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205

Qy        588 ---------------------------------------------------------- 588

Db        206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225

Qy        589 ------------CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 636
                          ||||||||||||||||   ||||||      :::   :::
Db        226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLeuValSerGlySerLeu 245
```

```
Qy        637 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 696
              |||:::|||   |||  |||  |||:::        :::          ||||||
Db        246 CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264

Qy        697 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 756
              ||||||:::   ||||||::::::|||   |||           ||||||  ||||||
Db        265 SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal 284

Qy        757 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 816
              |||::::::     |||::::::   :::      ::::::||||||:::|||   :::
Db        285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleIleLeu 303

Qy        817 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 876
              |||||||||||||||   ::::::|||||||||   |||      :::|||      |||
Db        304 HisSerTyrSerGlnLeuLeuMetPheProTyrGlyTyrLysCysThrLysLeuAspAsp 323

Qy        877 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 936
                 :::|||||||||   ||||||   :::|||   :::::::::   ::::::
Db        324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342

Qy        937 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 996
              |||:::|||      |||                :::|||   |||   ||||||
Db        343 GlyThrLysTyrLysValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer 362

Qy        997 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC------------ 1044
              ||||||   ||||||  ||||||||||||||||||
Db        363 IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr 382

Qy       1045 -----------------CCACCTGTAGAGAAGCTTTTGCCG 1068
                               |||      ::::::||||||
Db        383 GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro 396
```

RESULT 12
S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N;Alternate names: pancreatic carboxypeptidase A1
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: S29127; A34205; S08253; S02810; S71394; S02811
R;Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver
Biochem. J. 287, 299-303, 1992
A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.
A;Reference number: S29127; MUID:93038569; PMID:1417781
A;Accession: S29127
A;Molecule type: mRNA
A;Residues: 1-419 <CAT>
A;Cross-references: EMBL:X67318; NID:g35329; PIDN:CAA47732.1; PID:g35330
R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A;Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 330-396 <STE>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43;'XXX',114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

Alignment Scores:

Pred. No.:	1.18e-34	Length:	419
Score:	489.50	Matches:	131
Percent Similarity:	47.63%	Conservative:	60
Best Local Similarity:	32.67%	Mismatches:	127
Query Match:	25.34%	Indels:	83
DB:	1	Gaps:	13

US-09-980-881A-1_COPY_18_1097 (1-1080) x S29127 (1-419)

```
Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              |||      |||:::||||||:::      |||         |||
Db          4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluAsp------------ 19

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         ||||||||                              |||||||||   :::
Db         20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValGlnLysVal 39

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAG--------------- 165
              :::   |||              :::   |||      |||:::
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59

Qy        166 ------CCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 219
                   |||:::                          :::|||   |||
Db         60 GlySerProIleAspVal--------------------ArgValProPhe--------- 69

Qy        220 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 279
                     :::      ||||||      |||         ||||||
Db         70 ---ProSerIleGlnAlaValLysIlePheLeuGluSerHisGlyIleSerTyrGluThr 88

Qy        280 TTGCTGGCAGACGTGGAAGATCTTATT---------CAACAGCAGATT------------ 318
              ::::::   ||||||:::   |||:::        |||:::|||:::
Db         89 MetIleGluAspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaPheArg 108

Qy        319 ------------TCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
                         |||   ||||||   :::                        |||
Db        109 SerArgAlaArgSerThrAspThrPheAsn--------------------TyrAlaThr 121

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
              ||||||:::|||   ||||||||   :::::::::   :::         :::|||
Db        122 TyrHisThrLeuGluGluIleTyrAspPheLeuAspLeuLeuValAlaGluAsnProHis 141

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
              :::::::::||||||   ||||||:::::::::|||      |||:::||||||||||||
Db        142 LeuValSerLysIleGlnIleGlyAsnThrTyrGluGlyArgProIleTyrValLeuLys 161

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
                 |||            :::   :::   ||||||||||||||   ||||||||:::
Db        162 PheSer---ThrGlyGlySerLysArgProAlaIleTrpIleAspThrGlyIleHisSer 180

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
              |||||||||::::::   |||      :::||||||
Db        181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyr 200

Qy        585 ---------------------------------------------------------- 585
```

```
Db        201 GlyGlnAspAlaAlaPheThrAlaIleLeuAspThrLeuAspIlePheLeuGluIleVal 220

Qy        586 ----------------------------GGCCATAATCGAATGTGGAGAAAGAACCGT 615
                                          ||||||||||||||||||   |||
Db        221 ThrAsnProAspGlyPheAlaPheThrHisSerThrAsnArgMetTrpArgLysThrArg 240

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||      |||  :::  ||||||||  |||  |||  |||:::
Db        241 SerHisThrAlaGlySerLeuCysIleGlyValAspProAsnArgAsnTrp---AspAla 259

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
                 :::       ||||||||||||:::  ||||||||||||||  |||  :::
Db        260 GlyPheGlyLeuSerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyLysPhe 279

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
              ||||||  |||||||||:::::     |||:::::  :::  |||
Db        280 AlaAsnSerGluValGluValLysSerIleValAspPheValLysAspHisGlyAsn--- 298

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              |||||||||:::||||||:::|||||||||||||||  :::::::::||||||  |||
Db        299 IleLysAlaPheIleSerIleHisSerTyrSerGlnLeuLeuMetTyrProTyrGlyTyr 318

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
                          |||  :::||||||      ::::::      ||||||
Db        319 LysThrGluProValProAspGlnAspGluLeuAspGlnLeuSerLysAlaAlaValThr 338

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
              |||:::      :::          |||::::::  :::|||          :::  :::
Db        339 AlaLeu---AlaSerLeuTyrGlyThrLysPheAsnTyrGlySerIleIleLysAlaIle 357

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||  |||  |||          ||||||  |||      ||||||||||||||||||
Db        358 TyrGlnAlaSerGlySerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377

Qy       1036 ACA 1038
              |||
Db        378 Thr 378

RESULT 13
T33527
hypothetical protein T06A4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C;Accession: T33527
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33527
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-528 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.1
A;Map position: 1
A;Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C;Superfamily: carboxypeptidase

Alignment Scores:
Pred. No.:              4.22e-30        Length:         528
Score:                  437.50          Matches:        129
Percent Similarity:     45.28%          Conservative:   63
Best Local Similarity:  30.42%          Mismatches:     139
Query Match:            22.64%          Indels:         93
DB:                     2               Gaps:           16

US-09-980-881A-1_COPY_18_1097 (1-1080) x T33527 (1-528)

Qy          1 ATGAAGCTTTGCAGC---CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCAT 57
              |||      |||      ||||||  |||:::           :::      :::|||

Db          1 MetThrLysCysGlnIleLeuAlaSerLeuLeuThrTrpSerPheTrpValPheArgHis 20

Qy         58 GTCTTCGCGTTCCAGAGTGGC---------------------------------------- 78
              |||  |||       |||
Db         21 ValLeuAlaThrThrAspGlyArgAlaPheArgLysLysSerSerAlaThrThrProPro 40

Qy         79 --------------------CAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTT 117
                                   :::::::::         |||  |||         |||
Db         41 AsnGlnLysThrSerSerPheLysLeuIleArgIleAsnProGluThrGluGlySerVal 60

Qy        118 CAAGTTCTACAGAATCTT---------ACTACAACATATGAGATTGTTCTCTGGCAGCCG 168
              :::   |||::::::|||          :::  ||||||:::     |||||||||
Db         61 LysTyrLeuArgSerLeuTyrGluAspProSerProTyrGluLeuAspPheTrpGlnPro 80

Qy        169 GTAACA--------------GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTT 213
              |||                |||||  |||                      |||
Db         81 ProThrAsnIleGlyAlaIleValAspLeuThrValAlaProAlaAspAlaProArgPhe 100

Qy        214 GTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 273
              |||         |||:::::::::   |||                      |||
Db        101 Val------LysAspLeuGluSerLysLys--------------------IleSerTyr 111

Qy        274 AGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 333
                |||  :::  |||:::      |||:::  :::      :::|||
Db        112 IleValAlaValAsnAspLeuSerLysAlaIleGluAsnGluArgGlySerAspLysPhe 131

Qy        334 AGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCT 393
                     :::         |||::::::|||:::||||||   ||||||   :::
Db        132 TyrAsnProValAlaGlyPheAlaTyrAspLysTyrAsnSerLeuGluGluIleGlnThr 151

Qy        394 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 453
              ::::::  :::  :::  :::|||  |||:::|||  |||  ||||||
Db        152 GluMetLysArgLeuLysLysGluTyrProThrMetIleThrLeuIleAspIleGlyGln 171

Qy        454 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA------ 507
              |||  |||      |||  |||:::|||::::::||||||
Db        172 SerHisGluAsnArgThrLeuLeuValMetLysIleThrGlyLysArgAsnProLeuGly 191

Qy        508 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 567
              :::|||  ::::::|||||||||  |||||||||||||||||||||:::||||||
Db        192 SerLysIleSerMetTrpIleAspAlaGlyIleHisAlaArgGluTrpIleAlaProAla 211

Qy        568 TTCTGCTTGTGGTTCATAGGCCAT---------------------------------- 591
                  :::  :::|||  |||
Db        212 ThrAlaMet---TyrIleAlaHisGluLeuLeuLeuGlyTyrGluAsnAspAlaThrVal 230

Qy        591 ------------------------------------------------------------ 591

Db        231 AlaLysLeuMetAspHisIleAspPheTyrIleLeuProValMetAsnProAspGlyTyr 250

Qy        592 ------------------AATCGAATGTGGAGAAAGAACCGTTCT-------------- 618
                               |||||||||||||||||||||||||
Db        251 GluTyrSerArgGluLysAsnArgMetTrpArgLysAsnArgSerProAlaLysCysAla 270

Qy        619 ------TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
                    ::::::         |||  |||  |||||||||  ||||||
Db        271 ArgGlnThrPheSerThrValCysCysSerGlyValAspLeuAsnArgAsnPhe---Asp 289

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
              |||       |||:::||||||      |||  :::||||||  |||
Db        290 TrpPheTrpAlaSerThrGlySerSerSerAspProCysHisAspThrTyrHisGlySer 309

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                  ||||||||||||  :::||||||      ||||||  :::|||
Db        310 AlaAlaPheSerGluProGluSerGlnAlaValArgAspPheLeuGluGlnAsnThrPro 329

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              :::      |||:::||||||:::|||||||||||||||      :::   ||||||
Db        330 Glu------AlaPheIleSerLeuHisSerTyrSerGlnMetTrpLeuIleProTyrGly 347

Qy        853 TATACACGA---AGTAAAAGCAAAGAC---CATGAGGAACTGTCTCTAGTAGCCAGTGAA 906
              :::  :::  |||      :::|||  |||      |||      :::|||
Db        348 HisArgLysGlnSerTyrProGlnAspTyrHisThrGlyLeuArgProLeuAlaLeuArg 367
```

```
Qy        907 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 966
Db        368 AlaThrLysAlaLeuTyrGluLeu---TyrGlyThrLysTyrGlnValGlyThrGlyAla 386

Qy        967 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT---GATTTGGGCATC 1023
Db        387 AspLeuMetTyrGluAlaSerGlyGlySerHisAspTrpAlaLysGlyGlnLeuLysVal 406

Qy       1024 AAATATTCGTTT 1035
Db        407 ProTyrAlaTyr 410
```

RESULT 14
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20507; T24289
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284
A;Accession: T20507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WIL>
A;Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone F02D8
R;Wild, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19869
A;Accession: T24289
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WI2>
A;Cross-references: EMBL:Z78413; PIDN:CAB01666.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone T01C3
C;Genetics:
A;Gene: CESP:F02D8.4
A;Map position: 5
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C;Superfamily: carboxypeptidase

Alignment Scores:

Pred. No.:	1.42e-26	Length:	323
Score:	397.00	Matches:	88
Percent Similarity:	53.01%	Conservative:	53
Best Local Similarity:	33.08%	Mismatches:	81
Query Match:	20.55%	Indels:	44
DB:	2	Gaps:	7

US-09-980-881A-1_COPY_18_1097 (1-1080) x T20507 (1-323)

```
Qy        394 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 453
Db          3 PheLeuAsnSerLeuAlaGlnGlnTyrProAsnAspValLysLeuGlnAsnIleGlyAsn 22

Qy        454 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAA 513
Db         23 SerTyrGluGlyArgSerIleThrAlaValArgIleAla---AspAspGlySerSerLys 41

Qy        514 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGC 573
Db         42 ProIleValTrpIleAspAlaGlyIleHisAlaArgGluTrpIleSerTyrAsnValAla 61

Qy        574 TTGTGGTTCATA------------------------------------------------ 585
Db         62 LeuTyrLeuIleTyrThrIleValSerGlnProAlaTyrArgAsnLeuLeuAspSerVal 81

Qy        586 ----------------------------------------------------GGCCATAAT 594
Db         82 GlnLeuValValValProAsnThrAsnProAspGlyTyrGluTyrSerArgThrAsnAsp 101

Qy        595 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 654
Db        102 ArgMetTrpArgLysThrArgSerArgPheThrAsnSerArgCysAlaGlyAlaAspAla 121

Qy        655 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 714
Db        122 AsnArgAsnTyr---ProPheTyrTrpGlyThrGlnGlyValSerHisSerGlnCysSer 140

Qy        715 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 774
Db        141 GluIlePheCysGlySerArgProGlnSerGluProGluValLeuAlaLeuThrAsnAla 160

Qy        775 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 834
Db        161 IleIleArgAspGluGluArgIleLysGlyTyrIleAlaLeuHisSerTyrGlyGlnGlu 180

Qy        835 ATAGTGTTTCCATATTCCTATACA---CGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 891
Db        181 IleLeuTyrProTrpGlyHisThrGlnArgThrTyrProThrAspValGlnAspLeuIle 200

Qy        892 CTAGTAGCC---------AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 942
Db        201 GlnValGlyArgAlaMetAlaSerAlaIleArgAlaValAsn-----------AsnThr 216

Qy        943 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1002
Db        217 AspTyrThrValValAsnSerGlyAspGlyLeuTyrProAlaAlaGlyAlaSerAspAsp 236

Qy       1003 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC---CCACCTGTAGAGAAG 1059
Db        237 TrpAlaLysSerArgGlyIleLysTyrSerTyrThrIleGluLeuSerProIleAspAsp 256

Qy       1060 CTTTTGCCGCTGTCTCTA 1077
Db        257 PheThrGlyPheSerLeu 262
```

RESULT 15
T33526
hypothetical protein T06A4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33526
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-667 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67474.1; GSPDB:GN00019; CESP:T06A4.3
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.3
A;Map position: 1
A;Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

Alignment Scores:

Pred. No.:	2.01e-26	Length:	667
Score:	395.50	Matches:	121
Percent Similarity:	41.91%	Conservative:	63
Best Local Similarity:	27.56%	Mismatches:	138
Query Match:	20.47%	Indels:	117
DB:	2	Gaps:	15

US-09-980-881A-1_COPY_18_1097 (1-1080) x T33526 (1-667)

```
Qy         34 ATTGTTCTCTTCTGTGAGCAGCATGTC---TTCGCGTTCCAGAGTGGCCAAGTTCTAGCT 90
Db          8 LeuIlePheLeuCysPheSerAlaValSerLeuSerThrGlnSerPheAlaValTyrArg 27

Qy         91 GCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT---------ACTACA 141
```

```
Db         28 LeuLeuProLysSerGlnThrAspPheGlnAlaIleGlnArgLeuTyrLysAsnAlaThr 47

Qy        142 ACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCT--------------GACCTTATT 186
                 :::::::::      |||:::                          |||::::::
Db         48 AspHisAspLeuAsnPheTrpLysThrGlyLysAspLysHisGlyPheTrpAspValMet 67

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
               |||       |||         :::|||::::::
Db         68 ValAspMetLysAsnSerLysTyrPheLeuAsp-------------------------- 78

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
               |||   |||:::   |||              :::   |||||||||   ||||||
Db         79 PheLeuGlnValAsnAspIleSerTyrIleLysThrIleAspAspValGluGlyLeuIle 98

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGC---------CCCCGA-------------- 342
                  :::      :::|||   |||:::|||         ||||||
Db         99 ThrLysHisGluAsnAsnLysThrLeuSerAsnMetPheProArgLeuTrpAspAspSer 118

Qy        343 GCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATA 399
               :::||||||   |||   :::      |||   |||         :::       |||:::
Db        119 SerSerAlaHisTyrAspPheHisThrTyrGlySerTyrGlnArgMetThrAspTrpMet 138

Qy        400 GAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTT 459
               :::   :::      ::::::|||   |||:::      |||   ||||||   :::
Db        139 LysGlnLeuValValLysTyrProLysMetValGlnTyrIleSerIleGlyLysThrThr 158

Qy        460 GAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCC 519
               |||         :::      :::::::::   |||      :::|||   |||
Db        159 GluGlyArgAsnIleAspGlyValGluIleGlyGlyAspSerArgThr---LysLysIle 177

Qy        520 ATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGG 579
               |||||||||   |||||||||||||||||||   :::|||         |||:::
Db        178 PheTrpIleAspGlyGlyIleHisAlaArgGluTrpAlaAlaProHisThrAlaLeuPhe 197

Qy        580 TTCATA---------------------------------------------------- 585
               ||||||
Db        198 PheIleHisGlnValCysAspProAlaProSerGluLysSerValPheArgLeuValPro 217

Qy        585 ---------------------------------------------------------- 585

Db        218 LeuGlyLeuHisLysCysTrpGlnArgGlnLysLysGlyLysLeuAspAsnLysTrpLeu 237

Qy        585 ---------------------------------------------------------- 585

Db        238 ThrSerArgAlaAsnGluProGlyIleLysLysLeuLeuAsnGluIleThrPheValVal 257

Qy        586 ----------------------------------------------------GGCCATAAT 594
                                                                   |||
Db        258 ValProCysLeuAsnProAspGlyTyrGluPheThrArgSerSerThrAsnProHisVal 277

Qy        595 CGAATGTGGAGAAAGAACCGTTCT-----------------------TTCTATGCGAAC 630
               |||:::|||||||||||||||||                     :::      |||
Db        278 ArgLeuTrpArgLysAsnArgSerLysMetGlnCysArgLysAspIleTrpGlyArgAsn 297

Qy        631 AATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAA 690
                  |||   |||   |||||||||   ||||||         |||:::   |||
Db        298 ArgCysCysArgGlyValAspLeuAsnArgAsnPhe---AspPheHisPheArgGluSer 316

Qy        691 GGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCA 750
               |||   |||         |||||||||   |||   |||      |||   |||||||||
Db        317 GlyThrSerAspAspProCysSerGluIleTyrGlnGlyProSerProPheSerGluPro 336

Qy        751 GAAGTGAAGGCAGTGGCTAGTTTCTTG------AGAAGAAATATCAACCAGATTAAAGCA 804
               |||   |||||||||          |||      :::|||        :::      |||
Db        337 GluAlaLysAlaValArgAspAlaLeuLeuSerGlnArgTyrLysGlyArgThrAspAla 356

Qy        805 TACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA--- 861
               ||||||::::::|||:::|||||||||      :::   ||||||   :::   :::
Db        357 TyrIleThrLeuHisThrTyrSerGlnIleTrpIleHisProTyrGlyHisLysLysAsp 376

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
               :::          |||   ::::::|||      |||      :::|||   :::|||:::
Db        377 AlaTyrProGlyAspIleLysAspLeuTyrGluValGlyLysLysAlaAlaGlnAlaLeu 396

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                  |||              |||:::|||      |||   |||::::::|||||||||
Db        397 ---LysArgValTyrGlyThrLysTyrValValGlySerGlyAlaAspThrLeuTyrPro 415

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATC---TATGATTTGGGCATCAAATATTCGTTT 1035
                  |||   ||||||   :::||||||      ::::::      :::|||:::   :::
Db        416 AlaSerGlyGlySerGluAspTrpAlaLysHisGluAlaLysValLysPheValTyr 434
```

Search completed: January 1, 2004, 19:00:49
Job time : 39.1685 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 15:41:10 ; Search time 11.3984 Seconds
(without alignments)
8911.561 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 255726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135748_24941/app_query.fasta_1.2958
-DB=SwissProt_41 -QFMT=fastan -SUFFIX=rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_34_@runat_31122003_135748_24941 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	692	35.8	415	1	CBPB_RAT	P19223	rattus norv
2	641	33.2	417	1	CBPB_HUMAN	P15086	homo sapien
3	638	33.0	417	1	CBPC_MOUSE	P15089	mus musculu
4	635	32.9	417	1	CBPC_HUMAN	P15088	homo sapien
5	629.5	32.6	401	1	CBPB_PIG	P09955	sus scrofa
6	627.5	32.5	416	1	CBPB_CANFA	P55261	canis famil
7	581.5	30.1	306	1	CBPB_BOVIN	P00732	bos taurus
8	555.5	28.8	309	1	CBPC_RAT	P21961	rattus norv
9	553	28.6	417	1	CBP2_RAT	P19222	rattus norv
10	533.5	27.6	419	1	CBPA_BOVIN	P00730	bos taurus
11	529.5	27.4	419	1	CBP1_RAT	P00731	rattus norv
12	511	26.4	417	1	CPB2_HUMAN	P48052	homo sapien
13	489.5	25.3	419	1	CBP1_HUMAN	P15085	homo sapien
14	483	25.0	421	1	CBP4_HUMAN	Q9ui42	homo sapien
15	404.5	20.9	304	1	CBPZ_SIMVI	P42788	simulium vi
16	386	20.0	433	1	CBPA_ANOGA	O02350	anopheles g
17	369.5	19.1	303	1	CBPB_ASTFL	P04069	astacus flu
18	299.5	15.5	430	1	YHT2_YEAST	P38836	saccharomyc
19	277.5	14.4	424	1	CBPT_THEVU	P29068	thermoactin
20	271	14.0	451	1	CBPS_STRGR	P18143	streptomyce
21	251	13.0	434	1	CBPS_STRCP	P39041	streptomyce
22	118.5	6.1	707	1	ORC1_SCHPO	P54789	schizosacch
23	114.5	5.9	1380	1	CBPD_HUMAN	O75976	homo sapien
24	112.5	5.8	1377	1	CBPD_MOUSE	O89001	mus musculu
25	105	5.4	1255	1	PER2_HUMAN	O15055	homo sapien
26	103.5	5.4	722	1	CPXM_MOUSE	Q9z100	mus musculu
27	102	5.3	734	1	CPXM_HUMAN	Q96sm3	homo sapien
28	100	5.2	376	1	YQGT_BACSU	P54497	bacillus su
29	99.5	5.2	445	1	TPH1_CHICK	P70080	gallus gall
30	98.5	5.1	488	1	TY3H_ANGAN	O42091	anguilla an
31	97	5.0	663	1	TAZ1_SCHPO	P79005	schizosacch
32	95	4.9	444	1	TPH1_HUMAN	P17752	homo sapien
33	92	4.8	525	1	PUR9_BUCAI	P57143	b bifunctio
34	91.5	4.7	465	1	STHA_ECO57	Q8x727	escherichia
35	91.5	4.7	465	1	STHA_ECOLI	P27306	escherichia
36	91.5	4.7	3588	1	SRF1_BACSU	P27206	bacillus su
37	91	4.7	297	1	Y103_HUMAN	Q15006	homo sapien
38	91	4.7	444	1	TPH1_RABIT	P17290	oryctolagus
39	91	4.7	982	1	ENV_SFV3L	P27399	simian foam
40	90	4.7	1451	1	SPT6_YEAST	P23615	saccharomyc
41	89.5	4.6	682	1	CJBA_BACUH	O86170	bacillus th
42	88.5	4.6	319	1	UN30_CAEEL	P52906	caenorhabdi
43	88.5	4.6	388	1	RFC5_CAEEL	P34429	caenorhabdi
44	88.5	4.6	466	1	STHA_SALTY	Q8xfp5	salmonella
45	88	4.6	454	1	CBPH_LOPAM	P37892	lophius ame

ALIGNMENTS

RESULT 1
CBPB_RAT
ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M23959; AAA40872.1; -.
DR EMBL; M23947; AAA40872.1; JOINED.
DR EMBL; M23950; AAA40872.1; JOINED.
DR EMBL; M23952; AAA40872.1; JOINED.
DR EMBL; M23953; AAA40872.1; JOINED.
DR EMBL; M23954; AAA40872.1; JOINED.
DR PIR; A32129; A32129.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -.

```
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    13       BY SIMILARITY.
FT   PROPEP      14   108       ACTIVATION PEPTIDE.
FT   CHAIN      109   415       CARBOXYPEPTIDASE B.
FT   METAL      174   174       ZINC (BY SIMILARITY).
FT   METAL      177   177       ZINC (BY SIMILARITY).
FT   METAL      302   302       ZINC (BY SIMILARITY).
FT   ACT_SITE   376   376       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   171   184       BY SIMILARITY.
FT   DISULFID   243   266       BY SIMILARITY.
FT   DISULFID   257   271       BY SIMILARITY.
SQ   SEQUENCE   415 AA;  47515 MW;  8EA06CCADE30B6F2 CRC64;

Alignment Scores:
Pred. No.:                6.12e-54      Length:          415
Score:                    692.00        Matches:         150
Percent Similarity:       55.09%        Conservative:    61
Best Local Similarity:    39.16%        Mismatches:      126
Query Match:              35.82%        Indels:          46
DB:                       1             Gaps:            7

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_RAT (1-415)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||   |||             |||   |||   :::            |||:::|||
Db          1 MetLeuLeuLeuLeuAlaLeuValSerValAlaLeuAlaHisAlaSerGluGluHis--- 19

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
                 |||           :::|||                              |||
Db         20 ------PheAspGlyAsnArgValTyrArgValSerValHisGlyGluAspHisValAsn 37

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ::::::|||   |||       |||   ||||||      |||:::|||   :::|||
Db         38 LeuIleGlnGluLeuAlaAsnThrLysGluIleAspPheTrpLysProAspSerAlaThr 57

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
                 :::              |||   |||   |||   |||   ||||||   :::|||
Db         58 GlnValLysProLeuThrThrValAspPheHisValLysAlaGluAspValAlaAspVal 77

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              :::      |||      :::   :::            ||||||:::::::::|||   :::
Db         78 GluAsnPheLeuGluGluAsnGluValHisTyrGluValLeuIleSerAsnValArgAsn 97

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
                 ::::::   |||      :::   |||         |||||||||         |||
Db         98 AlaLeuGluSerGlnPheAspSerHisThr---------ArgAlaSerGlyHisSerTyr 114

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
                 :::|||:::            |||   :::||||||:::   :::         :::
Db        115 ThrLysTyrAsnLysTrpGluThrIleGluAlaTrpIleGlnGlnValAlaThrAspAsn 134

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||::::::|||:::      ||||||::::::||||||         :::||||||
Db        135 ProAspLeuValThrGlnSerValIleGlyThrThrPheGluGlyArgAsnMetTyrVal 154

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||:::   ||||||   :::      |||   ||||||:::|||||||||||||
Db        155 LeuLysIle---GlyLysThrArgProAsnLysProAlaIlePheIleAspCysGlyPhe 173

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------- 585
              |||||||||||||||||||||||||||||||   ||||||:::
Db        174 HisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArg 193

Qy        585 ------------------------------------------------------------ 585

Db        194 ThrTyrAsnGlnGluIleHisMetLysGlnLeuLeuAspGluLeuAspPheTyrValLeu 213

Qy        586 ----------------GGCCAT----------------AATCGAATGTGGAGAAAG 609
                             |||:::                :::|||||||||||||||
Db        214 ProValValAsnIleAspGlyTyrValTyrThrTrpThrLysAspArgMetTrpArgLys 233

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
                ||||||      |||   :::   |||:::|||         |||   ||||||
Db        234 ThrArgSerThrMetAlaGlySerSerCysLeuGlyValArgProAsnArgAsnPhe--- 252

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
                :::      |||||||||   |||||||||   |||   ||||||||||||||||||||
Db        253 AsnAlaGlyTrpCysGluValGlyAlaSerArgSerProCysSerGluThrTyrCysGly 272

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
                  ||||||||||||   |||   ||||||:::|||   |||:::|||   |||:::
Db        273 ProAlaProGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeu 292

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
                :::   ||||||||||||:::::::::|||||||||||||||   :::::::::||||||
Db        293 SerThrIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrProTyr 312

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
                ||||||                 :::::::::|||||||||::   :::         |||
Db        313 SerTyrAspTyrLysLeuProGluAsnTyrGluGluLeuAsnAlaLeuValLysGlyAla 332

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
                :::   :::         |||         |||:::||||||:::|||   |||:::
Db        333 AlaLysGluLeu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThr 351

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
                |||:::|||   |||   ||||||   |||||||||   ||||||   ||||||||||||
Db        352 ThrIleTyrProAlaAlaGlyGlySerAspAspTrpSerTyrAspGlnGlyIleLysTyr 371

Qy       1030 TCGTTTACA 1038
                |||||||||
Db        372 SerPheThr 374

RESULT 2
CBPB_HUMAN
ID   CBPB_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15086; O60834; Q96BQ8;
DT   01-APR-1990 (Rel. 14, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE   (PASP).
GN   CPB1 OR CPB OR PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC   TISSUE=Pancreas;
RX   MEDLINE=92129345; PubMed=1370825;
RA   Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA   French C.K.;
RT   "Isolation of a cDNA encoding a human serum marker for acute
RT   pancreatitis. Identification of pancreas-specific protein as
RT   pancreatic procarboxypeptidase B.";
RL   J. Biol. Chem. 267:2575-2581(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=98182241; PubMed=9524066;
RA   Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA   Aviles F.X.;
RT   "Comparative analysis of the sequences and three-dimensional models
RT   of human procarboxypeptidases A1, A2 and B.";
RL   Biol. Chem. 379:149-155(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
```

```
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 16-43.
RC   TISSUE=Pancreas;
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- TISSUE SPECIFICITY: Pancreas.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81057; AAA66973.1; -.
DR   EMBL; AJ224866; CAA12163.1; -.
DR   EMBL; BC015338; AAH15338.1; -.
DR   PDB; 1KWM; 28-AUG-02.
DR   MEROPS; M14.003; -.
DR   Genew; HGNC:2299; CPB1.
DR   MIM; 114852; -.
DR   GO; GO:0004180; F:carboxypeptidase activity; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1    15
FT   PROPEP      16   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   417       CARBOXYPEPTIDASE B.
FT   METAL      176   176       ZINC (BY SIMILARITY).
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173   186       BY SIMILARITY.
FT   DISULFID   245   268       BY SIMILARITY.
FT   DISULFID   259   273       BY SIMILARITY.
FT   CONFLICT    16    16       H -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    17    17       H -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    37    37       H -> Q (IN REF. 4).
FT   CONFLICT   208   208       D -> N (IN REF. 2).
FT   CONFLICT   245   245       MISSING (IN REF. 1).
SQ   SEQUENCE   417 AA;  47367 MW;  EBBB98B27F5D5AF9 CRC64;

Alignment Scores:
Pred. No.:              2.2e-49        Length:          417
Score:                  641.00         Matches:         140
Percent Similarity:     51.82%         Conservative:    73
Best Local Similarity:  34.06%         Mismatches:      137
Query Match:            33.18%         Indels:          62
DB:                     1              Gaps:            8

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_HUMAN (1-417)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              :::   :::|||   |||||||||   :::   |||           |||
Db          1 MetLeuAlaLeuLeuValLeuValThrValAlaLeuAlaSerAlaHisHisGlyGlyGlu 20

Qy         67 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
                 |||:::     :::|||                                 :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnIle 40

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 183
              ::::::  |||  :::|||  :::|||     |||:::|||  :::
Db         41 IleArgGluLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        184 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 243
              |||                |||  |||  |||  |||  |||         |||:::
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspThrValThrValGlu 80

Qy        244 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 303
                  |||     :::  :::         ||||||:::::::::::   ::::::
Db         81 AsnValLeuLysGlnAsnGluLeuGlnTyrLysValLeuIleSerAsnLeuArgAsnVal 100

Qy        304 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::  |||     :::             ||||||:::          ||||||
Db        101 ValGluAlaGlnPheAspSerArgVal---------ArgAlaThrGlyHisSerTyrGlu 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||:::           |||  :::|||  :::  :::         :::|||
Db        118 LysTyrAsnLysTrpGluThrIleGluAlaTrpThrGlnGlnValAlaThrGluAsnPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
                 ::::::::::       ||||||::::::||||||          :::|||:::|||
Db        138 AlaLeuIleSerArgSerValIleGlyThrThrPheGluGlyArgAlaIleTyrLeuLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              ||||||  ||||||       |||  ||||||:::::|||||||||   |||
Db        158 LysVal---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 585
              ||||||||||||||||||||||||||||||   ||||||::
Db        177 AlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaValArgThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyArgGluIleGlnValThrGluLeuLeuAspLysLeuAspPheTyrValLeuPro 216

Qy        586 --------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                           |||:::                 :::|||  |||||||||
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysSerArgPheTrpArgLysThr 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||   :::      :::  ||||||||||||||   |||   ||||||
Db        237 ArgSerThrHisThrGlySerSerCysIleGlyThrAspProAsnArgAsnPhe---Asp 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                  |||||||||   |||||||||   :::   |||   ||||||||||||||
Db        256 AlaGlyTrpCysGluIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyPro 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                  |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnLysLeuSer 295
```

```
Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              ||||||||||||:::::::::|||||||||||||||   :::::::::|||||||||
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetIleTyrProTyrSer 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              |||              :::::::::   ||||||:::   :::|||         |||
Db        316 TyrAlaTyrLysLeuGlyGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrVal 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
              :::   :::      :::         |||:::||||||:::|||   |||:::   |||
Db        336 LysGluLeu---AlaSerLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              :::|||   |||   ||||||   |||||||||   ||||||   ||||||:::||||||
Db        355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleArgTyrSer 374

Qy       1033 TTTACATC---------------------------------------------------- 1040
              ||||||
Db        375 PheThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              :::   ||||||   |||   |||   |||:::
Db        395 ArgAlaThrCysGluGluThrPheLeuAlaIle 405

RESULT 3
CBPC_MOUSE
ID   CBPC_MOUSE     STANDARD;      PRT;   417 AA.
AC   P15089;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90062123; PubMed=2584208;
RA   Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA   Serafin W.E.;
RT   "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT   novel member of the carboxypeptidase gene family.";
RL   J. Biol. Chem. 264:20094-20099(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05118; AAA37369.1; -.
DR   PIR; A34487; A34487.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   MGD; MGI:88479; Cpa3.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     15
FT   PROPEP      16    109       ACTIVATION PEPTIDE.
FT   CHAIN      110    417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL      176    176       ZINC (BY SIMILARITY).
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173    186       BY SIMILARITY.
FT   DISULFID   245    268       BY SIMILARITY.
SQ   SEQUENCE   417 AA;  48790 MW;  A2B300A068D1BA6D CRC64;

Alignment Scores:
Pred. No.:                4.07e-49      Length:          417
Score:                    638.00        Matches:         147
Percent Similarity:       51.23%        Conservative:    62
Best Local Similarity:    36.03%        Mismatches:      135
Query Match:              33.02%        Indels:          65
DB:                       1             Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPC_MOUSE (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGT 75
              :::||||||:::        :::          .  |||         |||
Db          7 MetAlaValIleTyrThrThrLeuAlaIleAlaProValHis---------PheAspArg 23

Qy         76 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 135
              :::|||                :::     :::          ||||||:::||||||
Db         24 GluLysValPheArgValLysLeuGlnAsnGluLysHisAlaSerValLeuLysAsnLeu 43

Qy        136 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAA 195
              |||   :::   |||:::       |||  |||                |||
Db         44 ThrGlnSerIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIleAlaValAsn 63

Qy        196 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 255
                 |||   |||   |||:::      :::         :::::::::   |||
Db         64 MetThrValAspPheArgValSerGluLysGluSerGlnThrIleGlnSerThrLeuGlu 83

Qy        256 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 315
                    |||         :::|||:::  |||:::::::::   |||::::::|||
Db         84 GlnHisLysIleHisTyrGluIleLeuIleHisAspLeuGlnGluGluIleGluLysGln 103

Qy        316 ATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCAC 372
                    |||   ::::::   |||   |||         |||   :::|||:::
Db        104 PheAspValLysAspGluIleAlaGlyArgHisSer---------TyrAlaLysTyrAsn 120

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
                 ::::::|||   ||||||   |||   :::   |||:::||||||:::|||:::
Db        121 AspTrpAspLysIleValSerTrpThrGluLysMetLeuGluLysHisProGluMetVal 140

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
              ::::::|||   |||||||||:::   |||       ||||||||||||||||||:::
Db        141 SerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLysIle--- 159

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
              ||||||:::        :::   ||||||::::::||||||||||||||||||||||
Db        160 GlyLysLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAlaArgGlu 179

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA--------------------GGCCAT 591
              ||||||||||||||||||||   ||||||:::                      |||
Db        180 TrpIleSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysSerTyrGlyLys 199

Qy        592 AAT---------------------------------------------------------- 594
              |||
Db        200 AsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProValPheAsn 219

Qy        595 --------------------------------CGAATGTGGAGAAAGAACCGTTCTTTC 621
                                             ||||||||||||||||||||||||
Db        220 ValAspGlyTyrIleTrpSerTrpThrGlnAspArgMetTrpArgLysAsnArgSerArg 239

Qy        622 TATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGG 681
                 |||::   ||||||||||||||||||||   ||||||           |||
Db        240 AsnGlnAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspValSerTrp 258

Qy        682 TGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAG 741
```

```
                          ::::::      |||         |||   |||      ||||||
Db        259 AspSerSerProAsnThrAsnLysProCysLeuAsnValTyrArgGlyProAlaProGlu 278

Qy        742 TCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAA 801
              ||||||   |||   |||||||||   :::|||:::|||   ::::::|||   ||||||
Db        279 SerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSerIleLys 298

Qy        802 GCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGA 861
              |||||||||:::   ||||||||||||||   ::::::   ||||||   ||||||
Db        299 AlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuIleProTyrGlyTyrThrPhe 318

Qy        862 AGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATT 921
                      :::|||::::::|||      ||||||      |||      |||:::
Db        319 LysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAspAlaLeu 338

Qy        922 GACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTA 981
                    |||         |||||||||   :::|||      :::   |||:::|||
Db        339 ---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIleTyrLys 357

Qy        982 GCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC---- 1037
                  |||         ||||||:::||||||||||||||||||:::::|||
Db        358 ThrSerGlySerSerLeuAspTrpValTyrAspLeuGlyIleLysHisThrPheAlaPhe 377

Qy       1038 ----------------------------------------------ATCAAACCCACC 1049
                                                            ||||||||||||
Db        378 GluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLysProThr 397

Qy       1050 TGTAGAGAAGCTTTTGCCGCTGTC 1073
              |||:::|||      :::|||
Db        398 CysLysGluThrMetLeuSerVal 405

RESULT 4
CBPC_HUMAN
ID   CBPC_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15088; Q96E94;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=90083291; PubMed=2594780;
RA   Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA   Austen K.F., Serafin W.E.;
RT   "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT   comparison of the protein with mouse mast cell carboxypeptidase A and
RT   rat pancreatic carboxypeptidases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mast cells;
RX   MEDLINE=92105393; PubMed=1729276;
RA   Reynolds D.S., Gurley D.S., Austen K.F.;
RT   "Cloning and characterization of the novel gene for mast cell
RT   carboxypeptidase A.";
RL   J. Clin. Invest. 89:273-282(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 110-417 FROM N.A.
RX   MEDLINE=92333165; PubMed=1629626;
RA   Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,
RA   Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT   "Human skin mast cell carboxypeptidase: functional characterization,
RT   cDNA cloning, and genealogy.";
RL   J. Invest. Dermatol. 99:138-145(1992).
RN   [5]
RP   SEQUENCE OF 110-137.
RX   MEDLINE=89214692; PubMed=2708524;
RA   Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT   "Human mast cell carboxypeptidase. Purification and
RT   characterization.";
RL   J. Clin. Invest. 83:1630-1636(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27717; AAA35652.1; -.
DR   EMBL; M73720; AAA59568.1; -.
DR   EMBL; M73716; AAA59568.1; JOINED.
DR   EMBL; M73717; AAA59568.1; JOINED.
DR   EMBL; M73718; AAA59568.1; JOINED.
DR   EMBL; M73719; AAA59568.1; JOINED.
DR   EMBL; BC012613; AAH12613.1; -.
DR   EMBL; S40234; AAB22578.2; ALT_SEQ.
DR   PIR; A43929; A43929.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   Genew; HGNC:2298; CPA3.
DR   MIM; 114851; -.
DR   GO; GO:0030141; C:secretory granule; NAS.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     15
FT   PROPEP      16    109       ACTIVATION PEPTIDE.
FT   CHAIN      110    417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL      176    176       ZINC (BY SIMILARITY).
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378    378       NUCLEOPHILE (BY SIMILARITY).
```

```
FT   DISULFID    173    186       BY SIMILARITY.
FT   DISULFID    245    268       BY SIMILARITY.
FT   CONFLICT     63     63       N -> K (IN REF. 3).
SQ   SEQUENCE   417 AA;  48700 MW;  8CB90DB758117B24 CRC64;

Alignment Scores:
Pred. No.:              7.54e-49       Length:         417
Score:                  635.00         Matches:        139
Percent Similarity:     52.31%         Conservative:   76
Best Local Similarity:  33.82%         Mismatches:     132
Query Match:            32.87%         Indels:         65
DB:                     1              Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPC_HUMAN (1-417)

Qy         16 CTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG--------- 66
              :::   :::::::::|||:::   |||                         |||
Db          1 MetArgLeuIleLeuProValGlyLeuIleAlaThrThrLeuAlaIleAlaProValArg 20

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         :::|||             |||:::     :::|||     ::::::
Db         21 PheAspArgGluLysValPheArgValLysProGlnAspGluLysGlnAlaAspIleIle 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              ::::::|||      |||   |||:::      |||   |||                ::: 
Db         41 LysAspLeuAlaLysThrAsnGluLeuAspPheTrpTyrProGlyAlaThrHisHisVal 60

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                          |||   |||   |||:::      :::         :::::::::
Db         61 AlaAlaAsnMetMetValAspPheArgValSerGluLysGluSerGlnAlaIleGlnSer 80

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
                 |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db         81 AlaLeuAspGlnAsnLysMetHisTyrGluIleLeuIleHisAspLeuGlnGluGluIle 100

Qy        307 CAACAGCAGATTTCC---AACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::|||           :::   :::      |||   |||         |||
Db        101 GluLysGlnPheAspValLysGluAspIleProGlyArgHisSer---------TyrAla 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db        118 LysTyrAsnAsnTrpGluLysIleValAlaTrpThrGluLysMetMetAspLysTyrPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::|||::::::::::|||   |||||||||:::   |||       ||||||||||||||
Db        138 GluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::   |||::::::          :::   ||||||::::::||||||||||||||
Db        158 LysIle---GlyGluLysAsnGluArgArgLysAlaIlePheMetAspCysGlyIleHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA----------------- 585
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        177 AlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaThrLysThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyArgAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrIleLeuPro 216

Qy        586 --------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                          |||:::                 ||||||||||||||||||||
Db        217 ValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAsnArgMetTrpArgLysAsn 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||         |||:::   ||||||||||||||||||||   ||||||   :::
Db        237 ArgSerLysAsnGlnAsnSerLysCysIleGlyThrAspLeuAsnArgAsnPhe---Asn 255

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
                  |||                 ::::::     |||::::::   |||   |||
Db        256 AlaSerTrpAsnSerIleProAsnThrAsnAspProCysAlaAspAsnTyrArgGlySer 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                 ||||||||||||   |||   |||||||||   :::|||:::|||   ::::::|||

Db        276 AlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsn 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        296 GluIleLysValTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGly 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
              ||||||                :::||||||:::|||:::   ||||||
Db        316 TyrThrSerLysLeuProProAsnHisGluAspLeuAlaLysValAlaLysIleGlyThr 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
                    :::       |||         |||||||||   :::|||             |||
Db        336 AspValLeu---SerThrArgTyrGluThrArgTyrIleTyrGlyProIleGluSerThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
              :::|||         |||     .   ||||||   ||||||||||||||||||:::::: 
Db        355 IleTyrProIleSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThr 374

Qy       1033 TTTAC----------------------------------------------------ATC 1040
              |||                                                     |||
Db        375 PheAlaPheGluLeuArgAspLysGlyLysPheGlyPheLeuLeuProGluSerArgIle 394

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              ||||||||||||||||          ||||||
Db        395 LysProThrCysArgGluThrMetLeuAlaVal 405

RESULT 5
CBPB_PIG
ID   CBPB_PIG       STANDARD;      PRT;   401 AA.
AC   P09955;
DT   01-MAR-1989 (Rel. 10, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2).
GN   CPB.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE OF 1-103.
RX   MEDLINE=91208150; PubMed=2018774;
RA   Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA   Aviles F.X.;
RT   "Analysis of the activation process of porcine procarboxypeptidase B
RT   and determination of the sequence of its activation segment.";
RL   Biochemistry 30:4082-4089(1991).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 1-38.
RX   MEDLINE=85279427; PubMed=4026847;
RA   Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT   "Sequential homologies between procarboxypeptidases A and B from
RT   porcine pancreas.";
RL   Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=91114690; PubMed=1989878;
RA   Coll M., Guasch A., Aviles F.X., Huber R.;
RT   "Three-dimensional structure of porcine procarboxypeptidase B: a
RT   structural basis of its inactivity.";
RL   EMBO J. 10:1-9(1991).
RN   [4]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX   MEDLINE=91027767; PubMed=2223783;
RA   Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT   "Sequence-specific 1H NMR assignments and determination of the
RT   secondary structure for the activation domain isolated from
RT   pancreatic procarboxypeptidase B.";
RL   Biochemistry 29:7515-7522(1990).
RN   [5]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=91114693; PubMed=1989879;
RA   Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
```

```
RT   "The NMR structure of the activation domain isolated from porcine
RT   procarboxypeptidase B.";
RL   EMBO J. 10:11-15(1991).
RN   [6]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=93044373; PubMed=1422143;
RA   Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA   Huber R., Wuethrich K.;
RT   "Comparison of the NMR solution structure with the X-ray crystal
RT   structure of the activation domain from procarboxypeptidase B.";
RL   J. Biomol. NMR 2:1-10(1992).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR   PDB; 1PBA; 31-OCT-93.
DR   PDB; 1NSA; 24-DEC-97.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW   3D-structure.
FT   PROPEP        1     95       ACTIVATION PEPTIDE.
FT   CHAIN        96    401       CARBOXYPEPTIDASE B.
FT   DISULFID    158    171
FT   DISULFID    230    253
FT   DISULFID    244    258
FT   METAL       161    161       ZINC.
FT   METAL       164    164       ZINC.
FT   METAL       289    289       ZINC.
FT   ACT_SITE    341    341
FT   ACT_SITE    363    363       NUCLEOPHILE.
FT   TURN          8      9
FT   STRAND       11     17
FT   HELIX        20     32
FT   STRAND       36     39
FT   HELIX        43     45
FT   STRAND       50     56
FT   HELIX        58     60
FT   HELIX        61     69
FT   TURN         70     72
FT   STRAND       74     79
FT   HELIX        82     87
FT   TURN         88     88
FT   TURN        101    102
FT   HELIX       107    120
FT   TURN        122    124
FT   STRAND      125    132
FT   TURN        134    135
FT   STRAND      138    144
FT   TURN        149    150
FT   STRAND      153    157
FT   TURN        162    163
FT   HELIX       165    181
FT   TURN        182    184
FT   HELIX       186    194
FT   STRAND      196    200
FT   HELIX       205    213
FT   TURN        214    214
FT   TURN        216    217
FT   STRAND      224    224
FT   TURN        226    227
FT   STRAND      231    231
FT   HELIX       235    237
FT   TURN        243    244
FT   STRAND      249    249
FT   TURN        252    253
FT   TURN        255    256
FT   STRAND      257    257
FT   TURN        262    263
FT   HELIX       266    277
FT   TURN        278    281
FT   STRAND      282    289
FT   STRAND      294    297
FT   TURN        307    308
FT   HELIX       309    327
FT   STRAND      332    335
FT   HELIX       336    339
FT   HELIX       347    353
FT   TURN        354    355
FT   STRAND      359    363
FT   TURN        371    372
FT   HELIX       376    378
FT   HELIX       379    398
FT   TURN        399    401
SQ   SEQUENCE   401 AA;  45713 MW;  53129AF159A26348 CRC64;

Alignment Scores:
Pred. No.:              2.32e-48      Length:          401
Score:                  629.50        Matches:         134
Percent Similarity:     54.85%        Conservative:    64
Best Local Similarity:  37.12%        Mismatches:      120
Query Match:            32.58%        Indels:          43
DB:                     1             Gaps:            6

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_PIG (1-401)

Qy        67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
             |||:::   :::|||                                 :::   |||
Db         7 PheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerGluLeu 26

Qy       127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
               |||  :::|||  :::|||      |||:::|||  :::       |||
Db        27 HisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGlnIle 46

Qy       187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
                        |||  |||  |||  |||  |||:::     |||:::
Db        47 LysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGluAsp 66

Qy       247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
               |||      :::  :::         ||||||:::  ::::::     ::::::
Db        67 PheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerValLeu 86

Qy       307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
             :::  |||     :::              |||  :::      ||||||:::
Db        87 GluAlaGlnPheAspSerArgCys---------ArgThrThrGlyHisSerTyrGluLys 103

Qy       367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
             |||::::::         |||  :::|||  |||  :::|||  ::::::||||||
Db       104 TyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValThrSerLysAsnProAsp 123

Qy       427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
             ::::::::::::      ||||||::::::|||:::         :::|||:::||||||
Db       124 LeuIleSerArgSerAlaIleGlyThrThrPheAspGlyAspAsnIleTyrLeuLeuLys 143

Qy       487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
             |||   ||||||      :::   |||   ||||||:::::|||||||||   ||||||
Db       144 Val---GlyLysProGlySerAsnLysProAlaIlePheMetAspCysGlyPheHisAla 162

Qy       547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
             ||||||||||||||   |||||||||   ||||||:::
Db       163 ArgGluTrpIleSerGlnAlaPheCysGlnTrpPheValArgAspAlaValArgThrTyr 182

Qy       585 ------------------------------------------------------------ 585

Db       183 GlyTyrGluAlaHisMetThrGluPheLeuAspAsnLeuAspPheTyrValLeuProVal 202

Qy       586 ------------GGCCAT------------------AATCGAATGTGGAGAAAGAACCGT 615
                         |||:::                  ||||||||||||||||||   |||
```

```
Db        203 LeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThrArg 222

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||      |||   :::   |||  |||||||||  |||  ||||||  :::
Db        223 SerThrAsnAlaGlySerSerCysThrGlyThrAspProAsnArgAsnPhe---AsnAla 241

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
              ||||||      |||||||||   :::   |||:::||||||||||||||
Db        242 GlyTrpCysThrValGlyAlaSerValAsnProCysAsnGluThrTyrCysGlySerAla 261

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
              |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        262 AlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSerSer 281

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              ||||||||||||:::::::::|||||||||||||||   |||::::::|||||||||||||
Db        282 IleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSerTyr 301

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
                  ::::::      ||||||:::   :::|||      ||||||:::
Db        302 AspTyrLysLeuProGluAsnAspAlaGluLeuAsnSerLeuAlaLysGlyAlaValLys 321

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
              :::      :::          |||   |||::::::|||   ||||||   |||:::
Db        322 GluLeu---AlaSerLeuTyrGlyThrSerTyrSerTyrGlyProGlySerThrThrIle 340

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||  |||   ||||||   |||||||||   |||:::   ||||||||||||||||||
Db        341 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAsnGlnGlyIleLysTyrSerPhe 360

Qy       1036 ACA 1038
              |||
Db        361 Thr 361

RESULT 6
CBPB_CANFA
ID   CBPB_CANFA     STANDARD;      PRT;   416 AA.
AC   P55261;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE   membrane associated protein) (ZAP47).
GN   CPB1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Fukuoka S.-I.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D78348; BAA11366.1; -.
DR   HSSP; P09955; 1PBA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    15       BY SIMILARITY.
FT   PROPEP      16   109       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN      110   416       CARBOXYPEPTIDASE B.
FT   METAL      175   175       ZINC (BY SIMILARITY).
FT   METAL      178   178       ZINC (BY SIMILARITY).
FT   METAL      303   303       ZINC (BY SIMILARITY).
FT   ACT_SITE   377   377       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   244   267       BY SIMILARITY.
FT   DISULFID   258   272       BY SIMILARITY.
SQ   SEQUENCE   416 AA;  47595 MW;  DFE1992CD52F8AB4 CRC64;

Alignment Scores:
Pred. No.:                3.52e-48        Length:          416
Score:                    627.50          Matches:         140
Percent Similarity:       51.10%          Conservative:    69
Best Local Similarity:    34.23%          Mismatches:      133
Query Match:              32.48%          Indels:          68
DB:                       1               Gaps:            9

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_CANFA (1-416)

Qy         22 GTCCTTGTACCCATTGTTCTC-----------TTCTGTGAGCAGCATGTCTTCGCGTTC 69
              :::||||||   :::  |||           :::     :::|||        |||
Db          5 IleLeuValThrLeuAlaLeuAlaSerAlaHisTyrSerGlyGluHis---------Phe 21

Qy         70 CAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAG 129
              :::      :::|||                                 :::   :::|||
Db         22 GluGlyGluLysValPheArgValAsnValGluAspGluAsnHisIleAsnLeuLeuHis 41

Qy        130 AATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTG 189
               |||  :::|||   :::|||      |||:::|||  :::         |||
Db         42 ThrLeuAlaSerThrThrGlnIleAspPheTrpLysProAspSerValThrGlnIleLys 61

Qy        190 AAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCAT 249
                        |||  |||  |||  |||:::     |||:::
Db         62 ProHisSerThrAlaAspPheArgValLysAlaGluAspIleLeuThrValGluAspPhe 81

Qy        250 TTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAA 309
              |||      :::   :::         ||||||:::   ::::::      :::::::::
Db         82 LeuLysGlnAsnGluLeuHisTyrGluValLeuIleAsnAsnLeuArgLeuValLeuGlu 101

Qy        310 CAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTAT 369
              |||                          |||:::         ||||||:::|||
Db        102 GlyGlnPheGlyArgGlnVal---------ProAlaThrGlyHisSerTyrGluLysTyr 118

Qy        370 CACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATG 429
              :::          |||  :::|||   :::   :::|||      :::||||||:::
Db        119 AsnArgTrpGluThrIleGluAlaTrpThrGlnGlnValThrSerGluAsnProAspLeu 138

Qy        430 CTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTT 489
              :::::::::      ||||||::::::||||||         :::|||:::|||||||||
Db        139 IleSerArgArgSerIleGlyThrThrPheGluGlyArgThrIleTyrLeuLeuLysVal 158

Qy        490 TCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGA 549
              ||||||              |||   ||||||:::::|||||||||   |||||||||
Db        159 ---GlyLysAlaGlyGlnAsnLysProAlaIlePheMetAspCysGlyPheHisAlaArg 177

Qy        550 GAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
              ||||||||||||||||||||      ||||||:::
Db        178 GluTrpIleSerProAlaPheTrpGlnTrpPheValArgGlu***IleArgThrTyrGly 197

Qy        585 ---------------------------------------------------------- 585

Db        198 GlnGluIleHisMetThrGluLeuLeuAspLysLeuAspPheTyrValLeuProValGly 217

Qy        586 ---------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCT 618
                      |||:::                 ||||||||||||||||||   ||||||
Db        218 AsnIleAspGlyTyrValTyrThrTrpThrLysAsnArgMetTrpArgLysThrArgSer 237
```

```
Qy        619 TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACAC 678
                          :::|||:::|||||||||          ||||||
Db        238 ThrGlnValGlyThrAsnCysValGlyThrAspProThrArgAsnPhe---AspAlaGly 256

Qy        679 TGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCT 738
              ||||||:::   |||||||||   :::   |||   ||||||||||||||||
Db        257 TrpCysLysIleGlyAlaSerArgAsnProCysAspGluThrTyrCysGlyProAlaAla 276

Qy        739 GAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATT 798
              |||||||||   |||   ||||||:::|||:::|||:::|||   |||::::::   |||
Db        277 GluSerGluLysGluThrLysAlaLeuAlaAsnPheIleArgSerAsnLeuSerSerIle 296

Qy        799 AAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACA 858
              |||||||||:::::::::|||||||||||||||   :::::::::|||||||||||||
Db        297 LysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrProTyrSerTyrAsp 316

Qy        859 CGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCT 918
                  ::::::::::::   ||||||:::   :::|||         |||:::
Db        317 TyrLysLeuThrGluAsnAsnAlaGluLeuAsnAlaLeuAlaLysAlaThrValLysGlu 336

Qy        919 ATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATAC 978
              :::      |||          |||:::||||||:::|||   |||:::   |||:::|||
Db        337 Leu---AlaThrLeuHisGlyThrLysTyrThrTyrGlyProGlyAlaThrThrIleTyr 355

Qy        979 CTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
                  |||   ||||||   |||||||||   ||||||   ||||||||||||||||||
Db        356 ProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrSerPheThr 375

Qy       1039 TC----------------------------------------------------AAACCC 1046
                                                                     |||
Db        376 PheGluLeuArgAspLysGlyArgTyrGlyPheAlaLeuProGluSerGlnIleSerPro 395

Qy       1047 ACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              ||||||   |||          |||:::
Db        396 ThrCysGluGluThrLeuLeuAlaIle 404
```

RESULT 7
CBPB_BOVIN
ID CBPB_BOVIN STANDARD; PRT; 306 AA.
AC P00732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B (EC 3.4.17.2).
GN CPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RX MEDLINE=75217824; PubMed=1057162;
RA Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT "Amino-acid sequence of bovine carboxypeptidase B.";
RL Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN [2]
RP SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX MEDLINE=74260705; PubMed=4833744;
RA Schmidt J.J., Hirs C.H.W.;
RT "Primary structure of bovine carboxypeptidase B. Inferences from the
RT locations of the half-cystines and identification of the active site
RT arginine.";
RL J. Biol. Chem. 249:3756-3764(1974).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX MEDLINE=76265065; PubMed=957425;
RA Schmid M.F., Herriott J.R.;
RT "Structure of carboxypeptidase B at 2.8-A resolution.";
RL J. Mol. Biol. 103:175-190(1976).
RN [4]
RP ACTIVE SITE.
RX MEDLINE=70007159; PubMed=5344132;
RA Plummer T.H. Jr.;
RT "Isolation and sequence of peptides at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 244:5246-5253(1969).
RN [5]
RP ACTIVE SITE.
RX MEDLINE=73061487; PubMed=4565668;
RA Kimmel M.T., Plummer T.H. Jr.;
RT "Identification of a glutamic acid at the active center of bovine
RT carboxypeptidase B.";
RL J. Biol. Chem. 247:7864-7869(1972).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A93797; CPBOB.
DR PDB; 1CPB; 30-SEP-83.
DR MEROPS; M14.003; -.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT DISULFID 63 76
FT DISULFID 135 158
FT DISULFID 149 163
FT METAL 66 66 ZINC.
FT METAL 69 69 ZINC.
FT METAL 194 194 ZINC.
FT ACT_SITE 246 246
FT ACT_SITE 268 268 NUCLEOPHILE.
SQ SEQUENCE 306 AA; 34612 MW; C329D2655C44A172 CRC64;

Alignment Scores:

Pred. No.:	4.3e-44	Length:	306
Score:	581.50	Matches:	120
Percent Similarity:	56.66%	Conservative:	46
Best Local Similarity:	40.96%	Mismatches:	70
Query Match:	30.10%	Indels:	58
DB:	1	Gaps:	6

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPB_BOVIN (1-306)

```
Qy        358 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 417
              ||||||:::|||::::::          |||   :::|||   |||   :::
Db          6 TyrGluLysTyrAsnAsnTrpGluThrIleGluAlaTrpThrGluGlnValAlaSerGlu 25

Qy        418 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 477
              :::||||||:::::::::::::      ||||||::::::|||            :::|||
Db         26 AsnProAspLeuIleSerArgSerAlaIleGlyThrThrPheLeuGlyAsnThrIleTyr 45

Qy        478 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 537
              :::||||||||||   ||||||      :::   |||   |||:::::::::|||||||||
Db         46 LeuLeuLysVal---GlyLysProGlySerAsnLysProAlaValPheMetAspCysGly 64

Qy        538 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 585
                 ||||||||||||||||||||||||||||||   ||||||:::
Db         65 PheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPheValArgGluAlaVal 84

Qy        585 ------------------------------------------------------------ 585

Db         85 ArgThrTyrGlyArgGluIleHisMetThrGluPheLeuAspLysLeuAspPheTyrVal 104

Qy        586 --------------------GGCCAT-----------------AATCGAATGTGGAGA 606
                                 |||:::                 ||||||||||||||
Db        105 LeuProValValAsnIleAspGlyTyrIleTyrThrTrpThrThrAsnArgMetTrpArg 124

Qy        607 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 666
              |||   ||||||      |||   :::   |||   ||||||||||||||   ||||||
Db        125 LysThrArgSerThrArgAlaGlySerSerCysThrGlyThrAspLeuAsnArgAsnPhe 144

Qy        667 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 726
                          ||||||      |||||||||::::::   ||||||||||||||||||
```

```
Db        145 ---AspAlaGlyTrpCysSerIleGlyAlaSerAsnAsnProCysSerGluThrTyrCys 163

Qy        727 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 786
              |||         |||||||||   |||   ||||||||||||   |||:::|||   :::
Db        164 GlySerAlaAlaGluSerGluLysGluSerLysAlaValAlaAspPheIleArgAsnHis 183

Qy        787 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 846
              ::::::   ||||||||||||:::::::::|||||||||||||||   :::::::::|||
Db        184 LeuSerSerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetMetLeuTyrPro 203

Qy        847 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 906
              ||||||||                 |||::::::   ||||||:::   :::|||
Db        204 TyrSerTyrAspTyrLysLeuProLysAsnAsnValGluLeuAsnThrLeuAlaLysGly 223

Qy        907 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 966
              ||||||:::   :::      :::         |||   |||::::::|||   |||:::
Db        224 AlaValLysLysLeu---AlaSerLeuHisGlyThrThrTyrSerTyrGlyProGlyAla 242

Qy        967 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1026
                 |||:::|||   |||   ||||||   ||||||||   ||||||   ||||||||||
Db        243 ThrThrIleTyrProAlaSerGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLys 262

Qy       1027 TATTCGTTTACATC-------------------------------------------- 1040
              ||||||||||||
Db        263 TyrSerPheThrPheGluLeuArgAspLysGlyArgTyrGlyPheValLeuProGluSer 282

Qy       1041 ------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                    :::|||||||||   |||         |||:::
Db        283 GlnIleGlnProThrCysGluGluThrMetLeuAlaIle 295
```

```
RESULT 8
CBPC_RAT
ID   CBPC_RAT       STANDARD;      PRT;   309 AA.
AC   P21961;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE   A3).
GN   CPA3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91105153; PubMed=1988052;
RA   Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA   Neurath H.;
RT   "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT   enzyme activity within mast cell granules.";
RL   Biochemistry 30:648-655(1991).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A38395; A38395.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   METAL        68     68       ZINC (BY SIMILARITY).
FT   METAL        71     71       ZINC (BY SIMILARITY).
FT   METAL       196    196       ZINC (BY SIMILARITY).
FT   ACT_SITE    248    248       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    270    270       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     65     78       BY SIMILARITY.
FT   DISULFID    137    160       BY SIMILARITY.
FT   VARIANT       1      1       MISSING (IN THE MAJOR FORM).
SQ   SEQUENCE   309 AA;  35786 MW;  20330FABC3EE83EF CRC64;
```

```
Alignment Scores:
Pred. No.:              9.03e-42        Length:         309
Score:                  555.50          Matches:        118
Percent Similarity:     54.27%          Conservative:   41
Best Local Similarity:  40.27%          Mismatches:     77
Query Match:            28.75%          Indels:         58
DB:                     1               Gaps:           6

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPC_RAT (1-309)

Qy        358 TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGG 417
              |||   :::|||:::       |||:::|||   ||||||   |||   :::   |||:::
Db          8 TyrAlaLysTyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLys 27

Qy        418 CATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTAT 477
              ||||||:::|||:::::::::|||   |||||||||:::   |||       |||||||||
Db         28 HisProGluMetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyr 47

Qy        478 GTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGA 537
              |||||||||:::   |||::::::         :::   ||||||::::::|||||||||
Db         48 ValLeuLysIle---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGly 66

Qy        538 ATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA------------ 585
              |||||||||||||||||:::|||||||||||||||   ||||||:::
Db         67 IleHisAlaArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAla 86

Qy        586 ---------GGCCATAAT---------------------------------------- 594
                        |||   |||
Db         87 LysSerTyrGlyLysAsnAsnIleMetThrLysLeuLeuAspArgMetAsnPheTyrVal 106

Qy        595 ------------------------------------------------CGAATGTGGAGA 606
                                                             ||||||||||||
Db        107 LeuProValPheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArg 126

Qy        607 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 666
              ||||||||||||        ::::::   ||||||||||||||||||||   ||||||
Db        127 LysAsnArgSerLysAsnProSerSerThrCysIleGlyThrAspLeuAsnArgAsnPhe 146

Qy        667 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 726
                      |||                 :::   :::   |||         |||
Db        147 ---AspValSerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArg 165

Qy        727 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 786
              |||       |||||||||||   |||   |||||||||   :::|||:::|||   :::
Db        166 GlyProAlaProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHis 185

Qy        787 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 846
              :::|||   |||||||||||||:::   |||||||||||||   ::::::||||||
Db        186 LeuAsnSerIleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPhePro 205

Qy        847 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 906
              |||   ||||||                 :::|||::::::|||       ||||||
Db        206 TyrGlyTyrThrIleLysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIle 225

Qy        907 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 966
              |||         :::      :::         |||||||||   :::|||       :::
Db        226 AlaThrAspValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAla 244

Qy        967 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1026
                 |||:::|||         |||         ||||||   ||||||||||||||||||
Db        245 SerThrIleTyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLys 264

Qy       1027 TATTCGTTTAC----------------------------------------------- 1037
              ::::::|||
Db        265 HisThrPheAlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSer 284

Qy       1038 ---ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                  ||||||||||||||:::|||         :::|||
Db        285 ArgIleLysProThrCysLysGluThrMetLeuSerVal 297
```

```
RESULT 9
CBP2_RAT
ID   CBP2_RAT       STANDARD;      PRT;   417 AA.
AC   P19222;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=89034323; PubMed=3182871;
RA   Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA   Graf M., Rutter W.J.;
RT   "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT   cloning, and evolutionary implications on substrate specificity in
RT   the carboxypeptidase gene family.";
RL   J. Biol. Chem. 263:17828-17836(1988).
RN   [2]
RP   SEQUENCE OF 131-143 FROM N.A.
RX   MEDLINE=95386501; PubMed=7657630;
RA   Normant E., Gros C., Schwartz J.C.;
RT   "Carboxypeptidase A isoforms produced by distinct genes or
RT   alternative splicing in brain and other extrapancreatic tissues.";
RL   J. Biol. Chem. 270:20543-20549(1995).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=92105124; PubMed=1761558;
RA   Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT   "Structural evolution of an enzyme specificity. The structure of rat
RT   carboxypeptidase A2 at 1.9-A resolution.";
RL   J. Biol. Chem. 266:24606-24612(1991).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23721; AAA40956.1; -.
DR   EMBL; M23714; AAA40956.1; JOINED.
DR   EMBL; M23715; AAA40956.1; JOINED.
DR   EMBL; M23716; AAA40956.1; JOINED.
DR   EMBL; M23717; AAA40956.1; JOINED.
DR   EMBL; M23718; AAA40956.1; JOINED.
DR   EMBL; M23719; AAA40956.1; JOINED.
DR   EMBL; M23720; AAA40956.1; JOINED.
DR   EMBL; S79837; -; NOT_ANNOTATED_CDS.
DR   PIR; A32128; A32128.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.002; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16       POTENTIAL.
FT   PROPEP      17    112       ACTIVATION PEPTIDE.
FT   CHAIN       113   417       CARBOXYPEPTIDASE A2.
FT   METAL       177   177       ZINC.
FT   METAL       180   180       ZINC.
FT   METAL       304   304       ZINC.
FT   ACT_SITE    378   378       NUCLEOPHILE.
FT   DISULFID    246   269       BY SIMILARITY.
FT   DISULFID    318   352       PROBABLE.
SQ   SEQUENCE   417 AA;  46912 MW;  BEDCC41A830F2D45 CRC64;

Alignment Scores:
Pred. No.:              1.58e-41      Length:         417
Score:                  553.00        Matches:        137
Percent Similarity:     50.37%        Conservative:   69
Best Local Similarity:  33.50%        Mismatches:     137
Query Match:            28.62%        Indels:         66
DB:                     1             Gaps:           13

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBP2_RAT (1-417)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||:::|||      |||         :::         ::::::|||::::::
Db          1 MetArgLeuThrLeuLeuLeuAlaAlaLeuLeuGlyTyrIleTyrCysGlnGlu------ 18

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
                 |||         |||||||||      :::|||           |||::::::
Db         19 ---ThrPheValGlyAspGlnValLeuGluIleIleProSerHisGluGluGlnIleArg 37

Qy        121 GTTCTACAGAATCTTACTACAACA------TATGAGATTGTTCTCTGGCAGCCGGTAACA 174
               |||     |||                 |||:::     |||:::     |||
Db         38 ThrLeuLeuGlnLeuGluAlaGluGluHisLeuGluLeuAspPheTrpLysSerProThr 57

Qy        175 GCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGAC 234
                     :::      :::   ||||||      |||      :::   :::
Db         58 ------------IleProGlyGluThrValHisValArgValProPheAlaSerIleGln 73

Qy        235 AATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTG 294
               ||||||      |||         ||||||      |||:::::::::   ||||||
Db         74 AlaValLysValPheLeuGluSerGlnGlyIleAspTyrSerIleMetIleGluAspVal 93

Qy        295 GAAGATCTTATTCAACAGCAGATTTCCAAC---------GACACAGTCAGCCCCCGAGCC 345
               :::   |||:::   |||:::              :::              |||
Db         94 GlnValLeuLeuAspGlnGluArgGluGluMetLeuPheAsnGlnGlnArgGluArgGly 113

Qy        346 TCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 405
                     :::|||   ||||||:::|||   |||||||||       ::::::
Db        114 GlyAsnPheAsnPheGluAlaTyrHisThrLeuGluGluIleTyrGlnGluMetAspAsn 133

Qy        406 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 465
               :::      :::|||   :::::::::|||:::::::::|||||||||||||||
Db        134 LeuValAlaGluAsnProGlyLeuValSerLysValAsnLeuGlySerSerPheGluAsn 153

Qy        466 TACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACAGCCAAAAATGCCATA 522
               |||:::   |||||||||   |||   |||   :::          |||   ||||||
Db        154 ArgProMetAsnValLeuLysPheSerThrGlyGlyAsp---------LysProAlaIle 170

Qy        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGG--- 579
               |||:::|||   |||||||||||||||||||:::::   |||       ||||||
Db        171 TrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAlaThrAlaLeuTrpThr 190

Qy        579 ------------------------------------------------------------ 579

Db        191 AlaAsnLysIleAlaSerAspTyrGlyThrAspProAlaIleThrSerLeuLeuAsnThr 210

Qy        580 ---------TTCATA--------------------GGCCAT----------------- 591
                      |||:::                    |||:::
Db        211 LeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrValPheSerGlnThrThr 230

Qy        592 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 651
               ||||||||||||||||||   ||||||      :::   :::   |||:::|||   |||
Db        231 AsnArgMetTrpArgLysThrArgSerLysArgSerGlySerGlyCysValGlyValAsp 250

Qy        652 CTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 711
                  |||   |||:::         ::::::         ||||||||||||||   |||
```

```
Db        251 ProAsnArgAsnTrp---AspAlaAsnPheGlyGlyProGlyAlaSerSerSerProCys 269

Qy        712 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 771
              |||::::::|||   |||      |||   ||||||   |||||||||::::::
Db        270 SerAspSerTyrHisGlyProLysProAsnSerGluValGluValLysSerIleValAsp 289

Qy        772 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 831
              |||::::::   :::      ::::::||||||:::|||::::::|||||||||||||||
Db        290 PheIleLysSerHis---GlyLysValLysAlaPheIleThrLeuHisSerTyrSerGln 308

Qy        832 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 891
                ::::::|||||||||   |||      :::|||      |||      ||||||
Db        309 LeuLeuMetPheProTyrGlyTyrLysCysThrLysProAspAspPheAsnGluLeuAsp 328

Qy        892 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACA 951
              ||||||   :::|||   :::|||:::   |||            |||   |||
Db        329 GluValAlaGlnLysAlaAlaGlnAlaLeu---LysArgLeuHisGlyThrSerTyrLys 347

Qy        952 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1011
              |||                :::|||   |||   ||||||      ||||||   |||
Db        348 ValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySerIleAspTrpAlaTyr 367

Qy       1012 GATTTGGGCATCAAATATTCGTTTACATCAAAC-------------------------- 1044
              |||||||||||||||||||||||
Db        368 AspLeuGlyIleLysTyrSerPheAlaPheGluLeuArgAspThrAlaPheTyrGlyPhe 387

Qy       1045 ---CCACCTGTAGAGAAGCTTTTGCCG 1068
                  |||   :::::::::||||||
Db        388 LeuLeuProAlaLysGlnIleLeuPro 396
```

```
RESULT 10
CBPA_BOVIN
ID   CBPA_BOVIN     STANDARD;      PRT;   419 AA.
AC   P00730;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A precursor (EC 3.4.17.1).
GN   CPA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91151335; PubMed=1998496;
RA   le Hueerou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
RT   "Cloning and nucleotide sequence of a bovine pancreatic
RT   preprocarboxypeptidase A cDNA.";
RL   Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=96096549; PubMed=8522204;
RA   Goo J.H., Kim K.H., Choi K.Y.;
RT   "Cloning, sequencing and expression of the gene encoding a major
RT   allotypic preprocarboxypeptidase A from bovine pancreas.";
RL   Gene 165:333-334(1995).
RN   [3]
RP   SEQUENCE OF 111-417.
RX   MEDLINE=71155183; PubMed=5102489;
RA   Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT   chymotryptic peptides of the cyanogen bromide fragment F-I.";
RL   Biochemistry 10:938-950(1971).
RN   [4]
RP   REVISIONS TO 138 AND 141.
RX   MEDLINE=72138789; PubMed=5143102;
RA   Petra P.H., Hermodson M.A., Walsh K.A., Neurath H.;
RT   "Characterization of bovine carboxypeptidase A (Allan).";
RL   Biochemistry 10:4023-4025(1971).
RN   [5]
RP   SEQUENCE OF 17-120.
RX   MEDLINE=89150306; PubMed=3147705;
RA   Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neurath H.;
RT   "The amino acid sequence of the activation peptide of bovine pro-
RT   carboxypeptidase A.";
RL   Biochimie 70:1137-1142(1988).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
RX   MEDLINE=83294519; PubMed=6887246;
RA   Rees D.C., Lewis M., Lipscomb W.N.;
RT   "Refined crystal structure of carboxypeptidase A at 1.54-A
RT   resolution.";
RL   J. Mol. Biol. 168:367-387(1983).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX   MEDLINE=96003618; PubMed=7556081;
RA   Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT   "The three-dimensional structure of the native ternary complex of
RT   bovine pancreatic procarboxypeptidase A with proproteinase E and
RT   chymotrypsinogen C.";
RL   EMBO J. 14:4387-4394(1995).
RN   [8]
RP   VARIANT ALLELIC.
RX   MEDLINE=69283620; PubMed=5817619;
RA   Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Identification of the amino acid replacements characterizing the
RT   allotypic forms of bovine carboxypeptidase A.";
RL   Biochemistry 8:2762-2768(1969).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC       COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC       PROPROTEINASE E.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M61851; AAA30426.1; -.
DR   EMBL; M61851; AAA30427.1; -.
DR   EMBL; Z33906; CAA83955.1; -.
DR   PIR; JN0126; CPBOA.
DR   PDB; 3CPA; 15-JAN-87.
DR   PDB; 4CPA; 22-OCT-84.
DR   PDB; 5CPA; 15-JAN-87.
DR   PDB; 6CPA; 15-OCT-91.
DR   PDB; 7CPA; 31-JAN-94.
DR   PDB; 8CPA; 31-JAN-94.
DR   PDB; 1CBX; 31-JAN-94.
DR   PDB; 1CPS; 15-OCT-94.
DR   PDB; 2CTB; 31-JAN-94.
DR   PDB; 2CTC; 31-JAN-94.
DR   PDB; 1ARL; 01-AUG-96.
DR   PDB; 1ARM; 17-AUG-96.
DR   PDB; 1BAV; 01-APR-97.
DR   PDB; 1YME; 12-FEB-97.
DR   PDB; 1CPX; 05-AUG-98.
DR   PDB; 1PYT; 27-JAN-97.
DR   PDB; 1ELL; 26-JUN-02.
DR   PDB; 1ELM; 26-JUN-02.
DR   PDB; 1F57; 06-SEP-00.
DR   PDB; 1HDQ; 15-NOV-01.
DR   PDB; 1HDU; 15-NOV-01.
DR   PDB; 1HEE; 23-NOV-01.
DR   PDB; 1M4L; 04-MAR-03.
DR   MEROPS; M14.001; -.
```

```
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure; Polymorphism.
FT   SIGNAL        1     16
FT   PROPEP       17    110       ACTIVATION PEPTIDE.
FT   CHAIN       111    419       CARBOXYPEPTIDASE A.
FT   METAL       179    179       ZINC.
FT   METAL       182    182       ZINC.
FT   METAL       306    306       ZINC.
FT   ACT_SITE    358    358       PROTON DONOR.
FT   ACT_SITE    380    380       NUCLEOPHILE.
FT   DISULFID    248    271
FT   VARIANT     289    289       I -> V (IN ALLELIC VARIANT).
FT   VARIANT     338    338       E -> A (IN ALLELIC VARIANT).
FT   VARIANT     415    415       L -> V (IN ALLELIC VARIANT).
FT   CONFLICT     95     95       S -> L (IN REF. 5).
FT   CONFLICT    199    199       D -> N (IN REF. 3).
FT   CONFLICT    203    203       D -> N (IN REF. 3).
FT   CONFLICT    224    224       D -> N (IN REF. 3).
FT   CONFLICT    232    232       Q -> E (IN REF. 3).
FT   CONFLICT    295    295       D -> N (IN REF. 3).
FT   TURN        114    116
FT   TURN        119    120
FT   HELIX       125    138
FT   TURN        140    142
FT   STRAND      143    150
FT   TURN        152    153
FT   STRAND      156    162
FT   STRAND      171    176
FT   TURN        180    181
FT   HELIX       183    199
FT   TURN        200    202
FT   HELIX       204    212
FT   STRAND      214    218
FT   HELIX       223    231
FT   TURN        232    232
FT   TURN        234    235
FT   STRAND      242    242
FT   TURN        244    245
FT   STRAND      249    249
FT   TURN        253    254
FT   TURN        261    262
FT   STRAND      267    267
FT   TURN        270    271
FT   TURN        273    274
FT   STRAND      275    275
FT   TURN        280    281
FT   HELIX       284    296
FT   STRAND      299    307
FT   STRAND      311    314
FT   TURN        324    325
FT   HELIX       326    344
FT   STRAND      349    352
FT   HELIX       353    356
FT   TURN        357    357
FT   HELIX       364    370
FT   TURN        371    372
FT   STRAND      375    381
FT   TURN        388    389
FT   HELIX       393    395
FT   HELIX       396    416
SQ   SEQUENCE   419 AA;  47082 MW;  21B86407B3BFC452 CRC64;

Alignment Scores:
Pred. No.:              8.73e-40        Length:         419
Score:                  533.50          Matches:        134
Percent Similarity:     49.87%          Conservative:   61
Best Local Similarity:  34.27%          Mismatches:     133
Query Match:            27.61%          Indels:         63
DB:                     1               Gaps:           11

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPA_BOVIN (1-419)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              |||      |||:::||||||:::         |||      |||
Db          4 LeuLeuIleLeuSerValLeuLeuGlyAlaAlaLeuGlyLysGluAsp------------ 19

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         |||||||||                              :::||||||   :::
Db         20 PheValGlyHisGlnValLeuArgIleThrAlaAlaAspGluAlaGluValGlnThrVal 39

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGG-----------CAGCCG 168
              :::   |||              :::   |||      |||           ||||||
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProGlyGlnPro 59

Qy        169 GTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGAT 228
                                                :::       |||
Db         60 ----------------------GlySerProIleAspValArgValProPheProSer 71

Qy        229 GTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCA 288
              :::      ||||||      |||         ||||||         :::::::::
Db         72 LeuGlnAlaValLysValPheLeuGluAlaHisGlyIleArgTyrArgIleMetIleGlu 91

Qy        289 GACGTGGAAGATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCC 339
              ||||||:::   |||:::         |||:::|||:::
Db         92 AspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaSerGlnSerArgAla 111

Qy        340 CGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 396
              |||::::::   ::::::   |||      ||||||:::|||:::|||||||||   :::
Db        112 ArgSerThrAsnThrPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrAspPhe 131

Qy        397 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 456
              ::::::   :::         ||||||   :::::::::|||:::   ||||||   |||
Db        132 MetAspLeuLeuValAlaGluHisProGlnLeuValSerLysLeuGlnIleGlyArgSer 151

Qy        457 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 516
              :::|||      |||:::||||||||||| |||          :::   :::
Db        152 TyrGluGlyArgProIleTyrValLeuLysPheSer---ThrGlyGlySerAsnArgPro 170

Qy        517 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 576
              |||||||||||||||   |||||||||:::||||||||||||:::   |||      :::
Db        171 AlaIleTrpIleAspLeuGlyIleHisSerArgGluTrpIleThrGlnAlaThrGlyVal 190

Qy        577 TGGTTCATA---------------------------------------------------- 585
              ||||||
Db        191 TrpPheAlaLysLysPheThrGluAspTyrGlyGlnAspProSerPheThrAlaIleLeu 210

Qy        585 ------------------------------------------------------------ 585

Db        211 AspSerMetAspIlePheLeuGluIleValThrAsnProAspGlyPheAlaPheThrHis 230

Qy        586 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 645
                    ||||||:::|||||||||   ||||||      :::::::::   |||:::|||
Db        231 SerGlnAsnArgLeuTrpArgLysThrArgSerValThrSerSerSerLeuCysValGly 250

Qy        646 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 705
                 |||   |||   |||:::         :::   :::   ||||||||||||||
Db        251 ValAspAlaAsnArgAsnTrp---AspAlaGlyPheGlyLysAlaGlyAlaSerSerSer 269

Qy        706 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 765
                 ||||||||||||||   |||   |||      ||||||   |||||||||:::::
Db        270 ProCysSerGluThrTyrHisGlyLysTyrAlaAsnSerGluValGluValLysSerIle 289

Qy        766 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 825
                 |||::::::   :::   |||      ||||||:::::|||:::|||||||||
Db        290 ValAspPheValLysAspHisGlyAsn---PheLysAlaPheLeuSerIleHisSerTyr 308

Qy        826 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 885
              ||||||   :::::::::||||||   ||||||               |||      |||
```

```
Db        309 SerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrGlnSerIleProAspLysThrGlu 328

Qy        886 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 945
              |||:::   ||||||     ||||||  |||:::  |||:::        |||
Db        329 LeuAsnGlnValAlaLysSerAlaValGluAlaLeu---LysSerLeuTyrGlyThrSer 347

Qy        946 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1005
              |||   :::|||            |||:::|||  |||  ||||||     ||||||
Db        348 TyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSerGlyGlySerIleAspTrp 367

Qy       1006 ATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
                |||:::  ||||||||||||||||||||
Db        368 SerTyrAsnGlnGlyIleLysTyrSerPheThr 378

RESULT 11
CBP1_RAT
ID   CBP1_RAT       STANDARD;      PRT;   419 AA.
AC   P00731;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82105986; PubMed=6275388;
RA   Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA   Pictet R.L., Valenzuela P., Rutter W.J.;
RT   "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT   characterization of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89034324; PubMed=3182872;
RA   Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT   "Structural characterization of the rat carboxypeptidase A1 and B
RT   genes. Comparative analysis of the rat carboxypeptidase gene
RT   family.";
RL   J. Biol. Chem. 263:17837-17845(1988).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01232; CAA24542.1; -.
DR   EMBL; J00713; AAA40893.1; -.
DR   EMBL; M23990; AAA40955.1; -.
DR   EMBL; M23960; AAA40955.1; JOINED.
DR   EMBL; M23985; AAA40955.1; JOINED.
DR   EMBL; M23986; AAA40955.1; JOINED.
DR   EMBL; M23987; AAA40955.1; JOINED.
DR   EMBL; M23988; AAA40955.1; JOINED.
DR   EMBL; M23989; AAA40955.1; JOINED.
DR   PIR; A00911; CPRTA.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16
FT   PROPEP      17    110       ACTIVATION PEPTIDE.
FT   CHAIN      111    419       CARBOXYPEPTIDASE A1.
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      182    182       ZINC (BY SIMILARITY).
FT   METAL      306    306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358    358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380    380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248    271       BY SIMILARITY.
FT   CONFLICT   196    196       I -> V (IN REF. 1).
FT   CONFLICT   261    263       FGM -> LGK (IN REF. 1).
FT   CONFLICT   347    347       K -> E (IN REF. 1).
SQ   SEQUENCE   419 AA;  47197 MW;  BB002D1CB99B7491 CRC64;

Alignment Scores:
Pred. No.:              1.99e-39      Length:         419
Score:                  529.50        Matches:        134
Percent Similarity:     50.77%        Conservative:   63
Best Local Similarity:  34.54%        Mismatches:     136
Query Match:            27.41%        Indels:         55
DB:                     1             Gaps:           10

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBP1_RAT (1-419)

Qy          4 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 63
              :::|||     |||::::::|||:::     |||         |||
Db          3 ArgLeuLeuIleLeuSerLeuLeuLeuGluAlaValCysGlyAsnGluAsn--------- 19

Qy         64 GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
              |||       |||||||||                         |||||||||
Db         20 ---PheValGlyHisGlnValLeuArgIleSerAlaAlaAspGluAlaGlnValGlnLys 38

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAGCCGGTAACAGCT 177
              ::::::  |||             :::  |||     |||:::
Db         39 ValLysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgAspAlaAlaArg 58

Qy        178 GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT 237
                 |||          :::|||  |||                 :::  :::
Db         59 AlaGlyIleProIleAspValArgValProPhe-----------ProSerIleGlnSer 74

Qy        238 GTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAA 297
              ||||||||  |||        |||||        :::::::::  |||||:::
Db         75 ValLysAlaPheLeuGluTyrHisGlyIleSerTyrGluIleMetIleGluAspValGln 94

Qy        298 GATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCC 348
              |||:::         :::||||||:::|||                      ::::::
Db         95 LeuLeuLeuAspGluGluLysGlnGlnMetSerAlaPheGlnAlaArgAlaLeuSerThr 114

Qy        349 GCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTT 405
              |||:::   |||      ||||||:::|||:::|||||||||  :::::::::
Db        115 AspSerPheAsnTyrAlaThrTyrHisThrLeuAspGluIleTyrGluPheMetAspLeu 134

Qy        406 ATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAG 465
              :::       ||||||  :::::::::||||||  ||||||::::::||||||
Db        135 LeuValAlaGluHisProGlnLeuValSerLysIleGlnIleGlyAsnThrPheGluGly 154

Qy        466 TACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGG 525
              |||::::::|||||||||  |||           |||   :::  |||||||||
Db        155 ArgProIleHisValLeuLysPheSer---ThrGlyGlyThrAsnArgProAlaIleTrp 173

Qy        526 ATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA 585
              ||||||  |||||||||:::|||||||||::::::   |||  .   :::||||||
Db        174 IleAspThrGlyIleHisSerArgGluTrpValThrGlnAlaSerGlyValTrpPheAla 193

Qy        586 GGCCAT------------------------------------------------------ 591

Db        194 LysLysIleThrLysAspTyrGlyGlnAspProThrPheThrAlaValLeuAspAsnMet 213

Qy        592 ----------------------------------------------------------AAT 594
                                                                        |||
```

```
Db        214 AspIlePheLeuGluIleValThrAsnProAspGlyPheAlaTyrThrHisLysThrAsn 233

Qy        595 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 654
              |||||||||||||||   ||||||           :::   |||:::|||   |||
Db        234 ArgMetTrpArgLysThrArgSerHisThrGlnGlySerLeuCysValGlyValAspPro 253

Qy        655 AATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 714
              |||   |||:::           :::         ||||||||||:::   ||||||
Db        254 AsnArgAsnTrp---AspAlaGlyPheGlyMetAlaGlyAlaSerSerAsnProCysSer 272

Qy        715 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 774
              |||||||||   |||   :::|||   ||||||   |||||||||:::::      |||
Db        273 GluThrTyrArgGlyLysPheProAsnSerGluValGluValLysSerIleValAspPhe 292

Qy        775 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 834
              :::      :::   |||   |||||||||:::||||||:::|||||||||||||||
Db        293 ValThrSerHisGlyAsn---IleLysAlaPheIleSerIleHisSerTyrSerGlnLeu 311

Qy        835 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 894
              :::::::::||||||   ||||||          :::   |||       ||||||
Db        312 LeuLeuTyrProTyrGlyTyrThrSerGluProAlaProAspGlnAlaGluLeuAspGln 331

Qy        895 GTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACAT 954
              :::|||      ||||||   |||:::      :::         |||:::::   :::
Db        332 LeuAlaLysSerAlaValThrAlaLeu---ThrSerLeuHisGlyThrLysPheLysTyr 350

Qy        955 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 1014
              |||         :::|||:::|||   |||   |||         ||||||   |||
Db        351 GlySerIleIleAspThrIleTyrGlnAlaSerGlySerThrIleAspTrpThrTyrSer 370

Qy       1015 TTGGGCATCAAATATTCGTTTACA 1038
              ||||||||||||||||||||||
Db        371 GlnGlyIleLysTyrSerPheThr 378
```

```
RESULT 12
CPB2_HUMAN
ID   CPB2_HUMAN     STANDARD;      PRT;   417 AA.
AC   P48052; Q96A12; Q96QN3;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC   TISSUE=Pancreas;
RX   MEDLINE=95204457; PubMed=7896805;
RA   Catasus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
RA   Billeter M.;
RT   "The sequence and conformation of human pancreatic
RT   procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT   three-dimensional model.";
RL   J. Biol. Chem. 270:6651-6657(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98111000; PubMed=9450539;
RA   Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,
RA   Aviles F.X.;
RT   "Characterisation and preliminary X-ray diffraction analysis of human
RT   pancreatic procarboxypeptidase A2.";
RL   FEBS Lett. 420:7-10(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98046021; PubMed=9384570;
RA   Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT   "The three-dimensional structure of human procarboxypeptidase A2.
RT   Deciphering the basis of the inhibition, activation and intrinsic
RT   activity of the zymogen.";
RL   EMBO J. 16:6906-6913(1997).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U19977; AAA74425.1; -.
DR   EMBL; BC007009; AAH07009.1; -.
DR   EMBL; BC014571; AAH14571.1; -.
DR   EMBL; BC015140; AAH15140.1; -.
DR   PIR; A56171; A56171.
DR   PDB; 1AYE; 13-JAN-99.
DR   PDB; 1DTD; 12-JUL-00.
DR   PDB; 1O6X; 30-JAN-03.
DR   MEROPS; M14.002; -.
DR   Genew; HGNC:2297; CPA2.
DR   MIM; 600688; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0007039; P:vacuolar protein catabolism; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17    112       ACTIVATION PEPTIDE.
FT   CHAIN       113    417       CARBOXYPEPTIDASE A2.
FT   METAL       177    177       ZINC.
FT   METAL       180    180       ZINC.
FT   METAL       304    304       ZINC.
FT   ACT_SITE    378    378       NUCLEOPHILE.
FT   DISULFID    246    269
FT   DISULFID    318    352
FT   CONFLICT     80     80       E -> G (IN REF. 2; AAH07009).
FT   CONFLICT    302    302       T -> I (IN REF. 1).
SQ   SEQUENCE   417 AA;  46828 MW;  3D9C073A98700A83 CRC64;
```

```
Alignment Scores:
Pred. No.:              8.91e-38        Length:         417
Score:                  511.00          Matches:        128
Percent Similarity:     48.73%          Conservative:   64
Best Local Similarity:  32.49%          Mismatches:     136
Query Match:            26.45%          Indels:         66
DB:                     1               Gaps:           11

US-09-980-881A-1_COPY_18_1097 (1-1080) x CPB2_HUMAN (1-417)

Qy         55 CATGTCTTC---------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACC 105
              |||::::::             |||          |||||||||      :::|||
Db         13 HisIleTyrCysLeuGluThrPheValGlyAspGlnValLeuGluIleValProSerAsn 32

Qy        106 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTC------ 159
                 |||::::::   |||      |||                 :::   |||
Db         33 GluGluGlnIleLysAsnLeuLeuGlnLeuGluAlaGlnGluHisLeuGlnLeuAspPhe 52

Qy        160 TGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 219
              |||:::      |||                 :::      |||      |||
Db         53 TrpLysSerProThr-----------ThrProGlyGluThrAlaHisValArgValPro 68

Qy        220 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 279
                 :::|||      ||||||      |||         ||||||      |||:::
Db         69 PheValAsnValGlnAlaValLysValPheLeuGluSerGlnGlyIleAlaTyrSerIle 88

Qy        280 TTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCC 339
              ::::::   ||||||:::   |||:::   ::::::                  :::
Db         89 MetIleGluAspValGlnValLeuLeuAspLysGluAsnGluGluMetLeuPheAsnArg 108

Qy        340 CGAGCCTCCGCATCG---------TACTATGAACAGTATCACTCACTAAATGAAATCTAT 390
              |||          |||            :::      ||||||:::|||    ||||||
Db        109 ArgArgGluArgSerGlyAsnPheAsnPheGlyAlaTyrHisThrLeuGluGluIleSer 128

Qy        391 TCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGA 450
                 ::::::   :::          ||||||   :::::::::|||::::::||||||
Db        129 GlnGluMetAspAsnLeuValAlaGluHisProGlyLeuValSerLysValAsnIleGly 148

Qy        451 TCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT---GGAAAAGAACAAACA 507
              ||||||||||||       |||:::   |||||||||   |||   |||   :::
Db        149 SerSerPheGluAsnArgProMetAsnValLeuLysPheSerThrGlyGlyAsp------ 166

Qy        508 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 567
                 |||    |||||||||:::|||    |||||||||||||||||||:::::    |||
Db        167 ---LysProAlaIleTrpLeuAspAlaGlyIleHisAlaArgGluTrpValThrGlnAla 185

Qy        568 TTCTGCTTGTGGTTCATAGGC---------------------------------------- 588
                 ||||||
Db        186 ThrAlaLeuTrpThrAlaAsnLysIleValSerAspTyrGlyLysAspProSerIleThr 205

Qy        588 ---------------------------------------------------------- 588

Db        206 SerIleLeuAspAlaLeuAspIlePheLeuLeuProValThrAsnProAspGlyTyrVal 225

Qy        589 ------------CATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCAT 636
                         ||||||||||||||||||    ||||||      :::   :::
Db        226 PheSerGlnThrLysAsnArgMetTrpArgLysThrArgSerLysValSerGlySerLeu 245

Qy        637 TGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCA 696
              |||:::|||   |||   |||   |||:::      :::                |||||
Db        246 CysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGlyProGlyAla 264

Qy        697 TCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTG 756
              ||||||:::   ||||||::::::|||   |||             ||||||   ||||||
Db        265 SerSerAsnProCysSerAspSerTyrHisGlyProSerAlaAsnSerGluValGluVal 284

Qy        757 AAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATG 816
              |||::::::      |||::::::   :::      ::::::||||||:::|||::::::
Db        285 LysSerIleValAspPheIleLysSerHis---GlyLysValLysAlaPheIleThrLeu 303

Qy        817 CATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGAC 876
              |||||||||||||   ::::::|||||||||   |||      :::|||      |||
Db        304 HisSerTyrSerGlnLeuLeuMetPheProTyrGlyTyrLysCysThrLysLeuAspAsp 323

Qy        877 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 936
                 :::|||||||||   ||||||   :::|||   :::::::::   ::::::
Db        324 PheAspGluLeuSerGluValAlaGlnLysAlaAlaGlnSerLeu---ArgSerLeuHis 342

Qy        937 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 996
                 |||:::|||      |||                 :::|||   |||   ||||||
Db        343 GlyThrLysTyrLysValGlyProIleCysSerValIleTyrGlnAlaSerGlyGlySer 362

Qy        997 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAAC------------ 1044
                 ||||||   ||||||   ||||||||||||||||||
Db        363 IleAspTrpSerTyrAspTyrGlyIleLysTyrSerPheAlaPheGluLeuArgAspThr 382

Qy       1045 ------------------CCACCTGTAGAGAAGCTTTTGCCG 1068
                                  |||      ::::::||||||
Db        383 GlyArgTyrGlyPheLeuLeuProAlaArgGlnIleLeuPro 396

RESULT 13
CBP1_HUMAN
ID   CBP1_HUMAN     STANDARD;      PRT;   419 AA.
AC   P15085; Q9BS67;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1 OR CPA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=93038569; PubMed=1417781;
RA   Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA   Wicker-Planquart C., Puigserver A.;
RT   "cDNA cloning and sequence analysis of human pancreatic
RT   procarboxypeptidase A1.";
RL   Biochem. J. 287:299-303(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 17-42.
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
```

```
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67318; CAA47732.1; -.
DR   EMBL; BC005279; AAH05279.1; -.
DR   PIR; S29127; S29127.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   Genew; HGNC:2296; CPA1.
DR   MIM; 114850; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A1.
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      182   182       ZINC (BY SIMILARITY).
FT   METAL      306   306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358   358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380   380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248   271       BY SIMILARITY.
FT   CONFLICT   276   276       H -> R (IN REF. 2).
SQ   SEQUENCE   419 AA;  47140 MW;  439FAFFFAEE958B1 CRC64;

Alignment Scores:
Pred. No.:                7.42e-36        Length:         419
Score:                    489.50          Matches:        131
Percent Similarity:       47.63%          Conservative:   60
Best Local Similarity:    32.67%          Mismatches:     127
Query Match:              25.34%          Indels:         83
DB:                       1               Gaps:           13

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBP1_HUMAN (1-419)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              |||      |||:::||||||:::      |||        |||
Db          4 LeuLeuValLeuSerValLeuLeuGlyAlaValPheGlyLysGluAsp------------ 19

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||      |||||||||                                 :::
Db         20 PheValGlyHisGlnValLeuArgIleSerValAlaAspGluAlaGlnValGlnLysVal 39

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAG--------------- 165
              :::   |||               :::   |||      |||:::
Db         40 LysGluLeuGluAspLeuGluHisLeuGlnLeuAspPheTrpArgGlyProAlaHisPro 59

Qy        166 ------CCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAAT 219
                   |||:::                    :::|||   |||
Db         60 GlySerProIleAspVal--------------------ArgValProPhe--------- 69

Qy        220 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 279
                  :::      ||||||      |||        ||||||
Db         70 ---ProSerIleGlnAlaValLysIlePheLeuGluSerHisGlyIleSerTyrGluThr 88

Qy        280 TTGCTGGCAGACGTGGAAGATCTTATT---------CAACAGCAGATT------------ 318
              ::::::   ||||||:::   |||:::         |||:::|||:::
Db         89 MetIleGluAspValGlnSerLeuLeuAspGluGluGlnGluGlnMetPheAlaPheArg 108

Qy        319 ------------TCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
                         |||   ||||||   :::                      |||
Db        109 SerArgAlaArgSerThrAspThrPheAsn--------------------TyrAlaThr 121

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
              ||||||:::|||   |||||||||   :::::::::   :::          :::|||
Db        122 TyrHisThrLeuGluGluIleTyrAspPheLeuAspLeuLeuValAlaGluAsnProHis 141

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
              :::::::::||||||   ||||||:::::::::|||      |||:::||||||||||||
Db        142 LeuValSerLysIleGlnIleGlyAsnThrTyrGluGlyArgProIleTyrValLeuLys 161

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
                 |||           :::   :::   ||||||||||||||   |||||||||:::
Db        162 PheSer---ThrGlyGlySerLysArgProAlaIleTrpIleAspThrGlyIleHisSer 180

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------------- 585
              |||||||||::::::   |||      :::||||||
Db        181 ArgGluTrpValThrGlnAlaSerGlyValTrpPheAlaLysLysIleThrGlnAspTyr 200

Qy        585 ---------------------------------------------------------- 585

Db        201 GlyGlnAspAlaAlaPheThrAlaIleLeuAspThrLeuAspIlePheLeuGluIleVal 220

Qy        586 -----------------------------GGCCATAATCGAATGTGGAGAAAGAACCGT 615
                                                |||||||||||||||||   |||
Db        221 ThrAsnProAspGlyPheAlaPheThrHisSerThrAsnArgMetTrpArgLysThrArg 240

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||      |||   :::   |||||||||   |||   |||   |||:::
Db        241 SerHisThrAlaGlySerLeuCysIleGlyValAspProAsnArgAsnTrp---AspAla 259

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
                 :::          |||||||||||:::   |||||||||||||||   |||   :::
Db        260 GlyPheGlyLeuSerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyLysPhe 279

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
                  ||||||   |||||||||:::::      |||::::::   :::   |||
Db        280 AlaAsnSerGluValGluValLysSerIleValAspPheValLysAspHisGlyAsn--- 298

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              |||||||||:::||||||:::||||||||||||||   :::::::::||||||   |||
Db        299 IleLysAlaPheIleSerIleHisSerTyrSerGlnLeuLeuMetTyrProTyrGlyTyr 318

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
                        |||   :::||||||      ::::::      ||||||
Db        319 LysThrGluProValProAspGlnAspGluLeuAspGlnLeuSerLysAlaAlaValThr 338

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
              |||:::      :::         |||::::::   :::|||         :::  :::
Db        339 AlaLeu---AlaSerLeuTyrGlyThrLysPheAsnTyrGlySerIleIleLysAlaIle 357

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||   |||   |||         ||||||   |||      ||||||||||||||||||
Db        358 TyrGlnAlaSerGlySerThrIleAspTrpThrTyrSerGlnGlyIleLysTyrSerPhe 377

Qy       1036 ACA 1038
              |||
Db        378 Thr 378

RESULT 14
CBP4_HUMAN
ID   CBP4_HUMAN     STANDARD;      PRT;   421 AA.
AC   Q9UI42;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN   CPA4 OR CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99310537; PubMed=10383164;
RA   Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT   "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT   deacetylase inhibitors during differentiation of prostate epithelial
RT   cancer cells.";
RL   Cancer Res. 59:2981-2988(1999).
RN   [2]
RP   IMPRINTING.
RX   MEDLINE=22483629; PubMed=12552318;
RA   Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA   Yoshiura K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA   Niikawa N., Kishino T.;
RT   "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT   imprinting domain.";
RL   Hum. Genet. 112:220-226(2003).
CC   -!- FUNCTION: Could be involved in the histone hyperacetylation
CC       pathway.
CC   -!- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,
CC       heart, intestine, kidney, liver and lung. Except for fetal brain
CC       that shows no imprinting, expression was found preferentially from
CC       the maternal allele.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095719; AAF23230.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.017; -.
DR   Genew; HGNC:15740; CPA4.
DR   MIM; 607635; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; NAS.
DR   GO; GO:0016573; P:histone acetylation; NAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17    113       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN       114    421       CARBOXYPEPTIDASE A4.
FT   METAL       181    181       ZINC (BY SIMILARITY).
FT   METAL       184    184       ZINC (BY SIMILARITY).
FT   METAL       308    308       ZINC (BY SIMILARITY).
FT   ACT_SITE    382    382       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    250    273       BY SIMILARITY.
SQ   SEQUENCE   421 AA;  47379 MW;  91883D25B705BC6A CRC64;

Alignment Scores:
Pred. No.:              2.82e-35        Length:         421
Score:                  483.00          Matches:        119
Percent Similarity:     48.92%          Conservative:   63
Best Local Similarity:  31.99%          Mismatches:     130
Query Match:            25.00%          Indels:         60
DB:                     1               Gaps:           10

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBP4_HUMAN (1-421)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         |||||||||           |||         ::::::      |||

Db         21 PhePheGlyAspGlnValLeuArgIleAsnValArgAsnGlyAspGluIleSerLysLeu 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTC------TGGCAGCCGGTAACAGCT--- 177
                 |||       :::       :::   |||       |||:::      ::::::

Db         41 SerGlnLeuValAsnSerAsnAsnLeuLysLeuAsnPheTrpLysSerProSerSerPhe 60

Qy        178 -----------GACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCT 225
                        |||::::::|||           :::       |||

Db         61 AsnArgProValAspValLeuValProSerValSerLeuGlnAlaPhe------------ 76

Qy        226 GATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTG 285
                        |||:::   |||       |||:::      :::|||   :::

Db         77 --------------LysSerPheLeuArgSerGlnGlyLeuGluTyrAlaValThrIle 91

Qy        286 GCAGACGTGGAAGATCTTATT---------CAACAGCAGATTTCCAACGACACAGTCAGC 336
                |||::::::   |||:::               ::::::   ::::::

Db         92 GluAspLeuGlnAlaLeuLeuAspAsnGluAspAspGluMetGlnHisAsnGluGlyGln 111

Qy        337 CCCCGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCT 393
                 |||:::|||   ::::::   |||       ||||||||||||      ||||||

Db        112 GluArgSerSerAsnAsnPheAsnTyrGlyAlaTyrHisSerLeuGluAlaIleTyrHis 131

Qy        394 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 453
                 ::::::   |||           ||||||:::      ::::::   ||||||

Db        132 GluMetAspAsnIleAlaAlaAspProAspLeuAlaArgArgValLysIleGlyHis 151

Qy        454 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAA 513
                 |||||||||       |||:::||||||||||||   |||       :::         :::

Db        152 SerPheGluAsnArgProMetTyrValLeuLysPheSerThrGlyLysGlyValArgArg 171

Qy        514 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGC 573
                 |||:::|||::::::   |||||||||:::||||||||||||||   |||

Db        172 ProAlaValTrpLeuAsnAlaGlyIleHisSerArgGluTrpIleSerGlnAlaThrAla 191

Qy        574 TTGTGG------------------------------------------------------ 579
                 :::|||

Db        192 IleTrpThrAlaArgLysIleValSerAspTyrGlnArgAspProAlaIleThrSerIle 211

Qy        580 ------------------TTCATA---------------------------------- 585
                               |||:::

Db        212 LeuGluLysMetAspIlePheLeuLeuProValAlaAsnProAspGlyTyrValTyrThr 231

Qy        586 ---GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 642
                    ||||||:::|||||||||   ||||||            :::   ||||||

Db        232 GlnThrGlnAsnArgLeuTrpArgLysThrArgSerArgAsnProGlySerSerCysIle 251

Qy        643 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 702
                 |||  |||   |||   |||:::   :::       :::       :::||||||||

Db        252 GlyAlaAspProAsnArgAsnTrp---AsnAlaSerPheAlaGlyLysGlyAlaSerAsp 270

Qy        703 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 762
                 :::  |||||||||   |||   |||   :::       ||||||   |||||||||:::

Db        271 AsnProCysSerGluValTyrHisGlyProHisAlaAsnSerGluValGluValLysSer 290

Qy        763 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 822
                 |||       |||::::::::::::   |||       |||   :::|||   :::||||||

Db        291 ValValAspPheIleGlnLysHisGlyAsn---PheLysGlyPheIleAspLeuHisSer 309

Qy        823 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 882
                 |||||||||   :::::::::||||||   |||:::       |||:::   |||   |||

Db        310 TyrSerGlnLeuLeuMetTyrProTyrGlyTyrSerValLysLysAlaProAspAlaGlu 329

Qy        883 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 942
                 ||||||       ||||||       |||   :::|||:::         |||       |||

Db        330 GluLeuAspLysValAlaArgLeuAlaAlaLysAlaLeuAlaSerValSer---GlyThr 348

Qy        943 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1002
                 |||       |||             |||:::|||   |||   |||           |||

Db        349 GluTyrGlnValGlyProThrCysThrThrValTyrProAlaSerGlySerSerIleAsp 368

Qy       1003 TGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
                 |||   ||||||   |||||||||::::::||||||

Db        369 TrpAlaTyrAspAsnGlyIleLysPheAlaPheThr 380
```

```
RESULT 15
CBPZ_SIMVI
ID   CBPZ_SIMVI     STANDARD;      PRT;   304 AA.
AC   P42788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS   Simulium vittatum (Black fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC   Simuliidae; Simulium.
OX   NCBI_TaxID=7192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Gut;
RX   MEDLINE=94093864; PubMed=8269093;
RA   Ramos A., Mahowald A., Jacobs-Lorena M.;
RT   "Gut-specific genes from the black fly Simulium vittatum encoding
RT   trypsin-like and carboxypeptidase-like proteins.";
RL   Insect Mol. Biol. 1:149-163(1993).
CC   -!- FUNCTION: Involved in the digestion of the blood meal.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Gut-specific.
CC   -!- INDUCTION: By blood meal.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08481; AAA18531.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.UPA; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   NON_TER       1      1
FT   METAL        58     58       ZINC (BY SIMILARITY).
FT   METAL        61     61       ZINC (BY SIMILARITY).
FT   METAL       184    184       ZINC (BY SIMILARITY).
FT   ACT_SITE    236    236       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    259    259       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    125    148       BY SIMILARITY.
SQ   SEQUENCE   304 AA;  34849 MW;  2E6E3FF8A6AA9144 CRC64;

Alignment Scores:
Pred. No.:              2.74e-28        Length:         304
Score:                  404.50          Matches:        93
Percent Similarity:     51.42%          Conservative:   52
Best Local Similarity:  32.98%          Mismatches:     86
Query Match:            20.94%          Indels:         51
DB:                     1               Gaps:           9

US-09-980-881A-1_COPY_18_1097 (1-1080) x CBPZ_SIMVI (1-304)

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              |||||||||:::|||   ||||||||||||||||:::::   :::   :::   ||||||
Db          1 GlnTyrHisThrLeuProGluIleTyrSerTrpLeuAspArgLeuValGlnGluHisPro 20

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::   :::     :::     |||   |||:::|||          :::      :::
Db         21 GluHisValGluProValValGlyGlyLysSerTyrGluGlyArgGluIleArgGlyVal 40

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||||||||   |||:::                  :::   ::::::      ||||||
Db         41 LysValSerTyrLysLys------GlyAsnProValValMetValGluSerAsnIleHis 58

Qy        544 GCCAGAGAATGGATCTCTCCTGCT---------TTCTGCTTG----------------- 576
              ||||||||||||||:::   |||         :::   |||
Db         59 AlaArgGluTrpIleThrAlaAlaThrThrThrTyrLeuLeuAsnGluLeuLeuThrSer 78

Qy        577 -------------------------------------TGGTTCATA------------ 585
                                                  |||:::|||
Db         79 LysAsnSerThrIleArgGluMetAlaGluAsnTyrAspTrpTyrIleProValThr 98

Qy        586 ---------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGTTCT 618
                      |||:::                 :::|||||||||||||||   ||||||
Db         99 AsnProAspGlyTyrValTyrThrHisThrThrAspArgMetTrpArgLysThrArgSer 118

Qy        619 TTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACAC 678
                    ::::::   |||   |||||||||   |||   |||:::   :::   |||
Db        119 ProAsnProAspSerLeuCysAlaGlyThrAspProAsnArgAsnTrp---AsnPheHis 137

Qy        679 TGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCT 738
              |||   |||:::|||   ||||||      |||:::|||||||||   |||
Db        138 TrpMetGluGlnGlyThrSerSerArgProCysThrGluThrTyrGlyGlyLysLysAla 157

Qy        739 GAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATT 798
                  ||||||   |||   ::::::   :::   ||||||:::                ||||||
Db        158 PheSerGluValGluThrArgSerPheSerAspPheLeuLysThrLeuLysGlyGlnIle 177

Qy        799 AAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACA 858
              |||   |||::::::   |||||||||||||||   ::::::|||||||||   :::|||
Db        178 LysValTyrLeuAlaPheHisSerTyrSerGlnLeuLeuLeuProTyrGlyHisThr 197

Qy        859 CGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCT 918
                    :::   :::|||:::::::|||      :::         |||   |||:::
Db        198 CysGlnHisThrTyrAsnHisAspAspLeuGlnAlaIleGlyAspAlaAlaAlaArgSer 217

Qy        919 ATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATAC 978
              :::   :::             |||   ||||||   |||:::      :::   :::|||
Db        218 LeuAlaGln---ArgTyrGlyThrAspTyrThrValGlyAsnIleTyrAspAlaIleTyr 236

Qy        979 CTAGCTCCTGGAGGTGGGGACGATTGGATCTAT---------GATTTGGGCATCAAATAT 1029
              |||   ||||||      ||||||   |||         |||:::   |||   |||
Db        237 ProAlaSerGlyGlySerMetAspTrpAlaTyrAspThrLeuAspIleProIleAlaTyr 256

Qy       1030 TCGTTTACATCAAACCCA-------------------------CCTGTAGAGAAGCTT 1062
              ::::::         |||                         |||      ::::::
Db        257 ThrTyrGluLeuArgProArgAspGlyTrpAsnGlyPheGlnLeuProAlaAsnGlnIle 276

Qy       1063 TTGCCG 1068
              :::|||
Db        277 IlePro 278


Search completed: January  1, 2004, 18:54:08
Job time : 30.3984 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: January 1, 2004, 18:10:36 ; Search time 54.1425 Seconds
(without alignments)
10294.941 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1932
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 1661050

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135748_24954/app_query.fasta_1.2958
-DB=SPTREMBL_23 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_190_@runat_31122003_135748_24954 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1884	97.5	360	4	Q9P2Y6	Q9p2y6 homo sapien
2	1835	95.0	423	4	Q15114	Q15114 homo sapien
3	1830	94.7	423	4	Q96IY4	Q96iy4 homo sapien
4	1550.5	80.3	422	11	Q9JHH6	Q9jhh6 mus musculu
5	1544.5	79.9	422	11	Q9QZF0	Q9qzf0 mus musculu
6	1522.5	78.8	422	11	Q9EQV9	Q9eqv9 rattus norv
7	1029	53.3	198	4	Q9NTI8	Q9nti8 homo sapien
8	706	36.5	416	13	Q9PUF2	Q9puf2 bothrops ja
9	688.5	35.6	408	13	Q8AXN3	Q8axn3 paralichthy
10	646	33.4	416	6	Q9XSP3	Q9xsp3 sus scrofa
11	630.5	32.6	412	11	P97597	P97597 rattus norv
12	611.5	31.7	437	4	Q8N4T0	Q8n4t0 homo sapien
13	610	31.6	434	13	Q8AVS2	Q8avs2 xenopus lae
14	581	30.1	419	13	Q8UUK1	Q8uuk1 gallus gall
15	564.5	29.2	374	4	Q8IVL8	Q8ivl8 homo sapien
16	554.5	28.7	436	11	Q8R4H4	Q8r4h4 mus musculu
17	529.5	27.4	419	13	Q8AXN4	Q8axn4 paralichthy
18	519.5	26.9	436	4	Q8WXQ8	Q8wxq8 homo sapien
19	516.5	26.7	419	6	Q9TV85	Q9tv85 sus scrofa
20	506.5	26.2	420	11	Q8BMK6	Q8bmk6 mus musculu
21	484.5	25.1	452	13	Q8UW45	Q8uw45 fugu rubrip
22	483	25.0	262	11	Q8BVD0	Q8bvd0 mus musculu
23	472	24.4	247	4	Q96KZ9	Q96kz9 homo sapien
24	460.5	23.8	419	13	Q8AXN5	Q8axn5 paralichthy
25	460	23.8	424	5	Q9VL87	Q9vl87 drosophila
26	460	23.8	428	5	Q8IGC4	Q8igc4 drosophila
27	459	23.8	1132	5	Q9W475	Q9w475 drosophila
28	451.5	23.4	422	5	Q9VLZ2	Q9vlz2 drosophila
29	440.5	22.8	226	4	Q8TDE8	Q8tde8 homo sapien
30	436	22.6	584	5	Q9TZH1	Q9tzh1 caenorhabdi
31	434.5	22.5	351	5	Q8T0Q1	Q8t0q1 drosophila
32	434	22.5	351	4	Q8NA08	Q8na08 homo sapien
33	418	21.6	445	5	Q9W478	Q9w478 drosophila
34	416	21.5	488	5	Q9BL88	Q9bl88 caenorhabdi
35	412.5	21.4	416	5	Q9VRZ3	Q9vrz3 drosophila
36	407.5	21.1	419	5	Q9VLZ1	Q9vlz1 drosophila
37	402	20.8	424	5	Q9VS66	Q9vs66 drosophila
38	401.5	20.8	430	5	Q8SZP6	Q8szp6 drosophila
39	400.5	20.7	430	5	Q9VL86	Q9vl86 drosophila
40	399.5	20.7	427	5	Q9U9K2	Q9u9k2 aedes aegyp
41	397	20.5	323	5	Q19121	Q19121 caenorhabdi
42	395.5	20.5	540	5	Q9TZH2	Q9tzh2 caenorhabdi
43	394	20.4	453	5	Q9V342	Q9v342 drosophila
44	385	19.9	423	5	O61532	O61532 drosophila
45	382.5	19.8	581	5	Q23318	Q23318 caenorhabdi

ALIGNMENTS

RESULT 1
Q9P2Y6
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).

```
DR   EMBL; AB011969; BAA90475.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   360 AA;  40935 MW;  BF670B2F7437C1CB CRC64;

Alignment Scores:
Pred. No.:              7.91e-181      Length:         360
Score:                  1884.00        Matches:        354
Percent Similarity:     98.89%         Conservative:   2
Best Local Similarity:  98.33%         Mismatches:     4
Query Match:            97.52%         Indels:         0
DB:                     4              Gaps:           0

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9P2Y6 (1-360)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnThrGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArg 220

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 AsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              ||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||
Db        301 SerGluAlaValArgAlaIleGluLysIleSerLysAsnThrArgTyrThrHisGlyHis 320

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340

Qy       1021 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 IleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360

RESULT 2
Q15114
ID   Q15114      PRELIMINARY;      PRT;   423 AA.
AC   Q15114;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PCPB protein.
GN   PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92042093; PubMed=1939207;
RA   Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT   "Isolation, molecular cloning, and partial characterization of a novel
RT   carboxypeptidase B from human plasma.";
RL   J. Biol. Chem. 266:21833-21838(1991).
DR   EMBL; M75106; AAA60042.1; -.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.009; -.
DR   Genew; HGNC:2300; CPB2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
FT   CHAIN        23    423       PLASMA CARBOXYPEPTIDASE B.
SQ   SEQUENCE   423 AA;  48442 MW;  9B383272F6EE79F4 CRC64;

Alignment Scores:
Pred. No.:              7.06e-176      Length:         423
Score:                  1835.00        Matches:        356
Percent Similarity:     86.23%         Conservative:   1
Best Local Similarity:  85.99%         Mismatches:     3
Query Match:            94.98%         Indels:         55
DB:                     4              Gaps:           2

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q15114 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
```

```
Db        41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy       181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
Db        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
Db        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
Db       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
Db       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
Db       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
Db       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy       541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
Db       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy       591 ------------------------------------------------------------ 591

Db       201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy       592 ------------------------------------------AATCGAATGTGGAGAAAG 609
Db       221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy       610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
Db       241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy       670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db       261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
Db       281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
Db       301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 320

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db       321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
Db       341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
Db       361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy      1030 TCGTT-------------------------------------------------------TAC 1037
Db       381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy      1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
```

```
Db       401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414

RESULT 3
Q96IY4
ID   Q96IY4      PRELIMINARY;      PRT;   423 AA.
AC   Q96IY4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B2 (Plasma).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007057; AAH07057.1; -.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   423 AA;  48412 MW;  9B383E03F6EE7CF5 CRC64;

Alignment Scores:
Pred. No.:              2.25e-175      Length:          423
Score:                  1830.00        Matches:         355
Percent Similarity:     85.99%         Conservative:    1
Best Local Similarity:  85.75%         Mismatches:      4
Query Match:            94.72%         Indels:          55
DB:                     4              Gaps:            2

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q96IY4 (1-423)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
```

```
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisIleThrGln 200

Qy        591 ------------------------------------------------------------ 591

Db        201 PheTyrGlyIleIleGlyGlnTyrThrAsnLeuLeuArgLeuValAspPheTyrValMet 220

Qy        592 -------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        221 ProValValAsnValAspGlyTyrAspTyrSerTrpLysLysAsnArgMetTrpArgLys 240

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        241 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAla 260

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        261 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 280

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 300

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 320

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        321 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 340

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        341 ValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 360

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 380

Qy       1030 TCGTT-------------------------------------------------------TAC 1037
              |||                                                       |||
Db        381 SerPheThrIleGluLeuArgAspThrGlyThrTyrGlyPheLeuLeuProGluArgTyr 400

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              |||||||||||||||||||||||||||||||||||||||||||
Db        401 IleLysProThrCysArgGluAlaPheAlaAlaValSerLys 414
```

RESULT 4
Q9JHH6
ID Q9JHH6 PRELIMINARY; PRT; 422 AA.
AC Q9JHH6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE (1110032P04Rik protein).
GN CPB2 OR TAFI OR 1110032P04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20341711; PubMed=10878383;
RA Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA Campbell W., Okada H.;
RT "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT whereas carboxypeptidase N is not.";
RL J. Immunol. 165:1053-1058(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20201996; PubMed=10739389;
RA Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT "Characterization of mouse thrombin-activatable fibrinolysis
RT inhibitor.";
RL Thromb. Haemost. 83:297-303(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AB021968; BAB03402.1; -.
DR EMBL; AF164524; AAF62385.1; -.
DR EMBL; AK004045; BAB23141.1; -.
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR MGD; MGI:1891837; Cpb2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 422 AA; 48870 MW; 99113755669D55CB CRC64;

Alignment Scores:

Pred. No.:	3.2e-147	Length:	422
Score:	1550.50	Matches:	302
Percent Similarity:	78.99%	Conservative:	25
Best Local Similarity:	72.95%	Mismatches:	32
Query Match:	80.25%	Indels:	56
DB:	11	Gaps:	3

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9JHH6 (1-422)

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||       |||   :::||||||   |||:::|||:::   |||||||||
Db          1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db         20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
                  |||   |||||||||:::|||||||||||||||||||||||||||||||:::|||
Db         60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
```

```
Db        80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
             ||||||:::||||||     ||||||||||||||||||||||||||||||||||||||||
Db       100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119

Qy       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
             ||||||||||||||||||||||||||||||||||||||||   ||||||||:::|||
Db       120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
             ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db       140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159

Qy       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
             ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db       160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy       541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
             ||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db       180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy       591 ------------------------------------------------------------ 591

Db       200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy       592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db       220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy       610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
             |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db       240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy       670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
             ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 279

Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
             ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db       280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
             :::   ||||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db       300 AspHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
             ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
             ||||||||||||:::      :::|||||||||||||||||||||   |||||||||
Db       340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
             :::|||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db       360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy      1030 TCGTT-------------------------------------------------------TAC 1037
             |||                                                         |||
Db       380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399

Qy      1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
             ||||||||||||||   ||||||   ||||||:::||||||
Db       400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLys 413

RESULT 5
Q9QZF0
ID   Q9QZF0      PRELIMINARY;      PRT;   422 AA.
AC   Q9QZF0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase U.
GN   CPB2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/CBAF1J;
RA   He Y.C., Broze G.;
RT   "Isolation and characterization of mouse liver carboxypeptidase B
RT   gene.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF186188; AAF00528.1; -.
DR   HSSP; P00730; 5CPA.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48832 MW;  B53FFFB09943954E CRC64;

Alignment Scores:
Pred. No.:              1.29e-146     Length:         422
Score:                  1544.50       Matches:        301
Percent Similarity:     78.74%        Conservative:   25
Best Local Similarity:  72.71%        Mismatches:     33
Query Match:            79.94%        Indels:         56
DB:                     11            Gaps:           3

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9QZF0 (1-422)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||       |||   :::||||||   |||:::|||:::   |||||||||
Db          1 MetLysLeuHisGlyLeuGlyIleLeuValAlaIleIleLeuTyr---GluGlnHisGly 19

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db         20 PheAlaPheGlnSerGlyGlnValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              |||   |||||||||:::|||||||||||||||||||||||||||||||||||:::|||
Db         60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAspSerVal 79

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
Db         80 LysAlaHisLeuAsnValSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||:::||||||     ||||||||||||||||||||||||||||||||||||||||
Db        100 LeuIleGluGlnGlnThrPheAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 119

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||   :::||||||:::|||
Db        120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValLeuThrGluGlnHis 139

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        140 ProAspMetLeuGlnLysIleTyrIleGlySerSerPheGluLysTyrProLeuTyrVal 159

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db        160 LeuLysValSerGlyLysGluGlnArgIleLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::
```

```
Db        180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy        591 ------------------------------------------------------------ 591

Db        200 PheHisGlyLysGluAsnLeuTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        240 AsnArgSerAlaHisLysAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        260 SerLysHisTrpCysGluLysGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 279

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              :::   ||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        300 AspHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuPheProTyr 319

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||         :::||||||||||||||||||||||||   |||||||||
Db        340 ValArgAlaIleGlySerPheAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              :::||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy       1030 TCGTT----------------------------------------------------TAC 1037
              |||                                                     |||
Db        380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgTyr 399

Qy       1038 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
              ||||||||||||||   ||||||   ||||||:::||||||
Db        400 IleLysProThrCysAlaGluAlaLeuAlaAlaIleSerLys 413

RESULT 6
Q9EQV9
ID   Q9EQV9      PRELIMINARY;      PRT;   422 AA.
AC   Q9EQV9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Pre-procarboxypeptidase R.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20471387; PubMed=11021404;
RA   Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA   Okada N., Okada H.;
RT   "Molecular cloning and partial characterization of rat
RT   procarboxypeptidase R and carboxypeptidase N.";
RL   Microbiol. Immunol. 44:719-728(2000).
DR   EMBL; AB042598; BAB18617.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   422 AA;  48826 MW;  FFFD32A51A9366C8 CRC64;

Alignment Scores:
Pred. No.:              2.12e-144      Length:          422
Score:                  1522.50        Matches:         297
Percent Similarity:     78.02%         Conservative:    26
Best Local Similarity:  71.74%         Mismatches:      36
Query Match:            78.80%         Indels:          56
DB:                     11             Gaps:            3

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9EQV9 (1-422)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||      |||   |||||||||   |||:::|||:::   |||:::|||
Db          1 MetLysLeuTyrGlyLeuGlyValLeuValAlaIleIleLeuTyr---GluLysHisGly 19

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
                |||||||||||||||   ||||||:::|||||||||||||||||||||||||||||||
Db         20 LeuAlaPheGlnSerGlyHisValLeuSerAlaLeuProArgThrSerArgGlnValGln 39

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              :::|||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         40 LeuLeuGlnAsnLeuThrThrThrTyrGluValValLeuTrpGlnProValThrAlaGlu 59

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
                |||   |||||||||:::|||||||||||||||||||||||||||:::::|||
Db         60 PheIleGluLysLysLysGluValHisPhePheValAsnAlaSerAspValAsnSerVal 79

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||:::||||||   |||   ||||||  :::|||||:::   :::|||||||||
Db         80 LysAlaTyrLeuAsnAlaSerArgIleProPheAsnValLeuMetAsnAsnValGluAsp 99

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||   |||||||||||||||||||||||||:::|||||||||
Db        100 LeuIleGlnGlnGlnThrSerAsnAspThrValSerProArgAlaSerSerSerTyrTyr 119

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db        120 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluValIleThrGluGlnHis 139

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||   ||||||:::|||||||||||:::|||||||||||||||||||||
Db        140 ProAspMetLeuGlnLysIleTyrIleGlySerSerTyrGluLysTyrProLeuTyrVal 159

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        160 LeuLysValSerGlyLysGluHisArgValLysAsnAlaIleTrpIleAspCysGlyIle 179

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        180 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyTyrValThrGln 199

Qy        591 ------------------------------------------------------------ 591

Db        200 PheHisGlyLysGluAsnThrTyrThrArgLeuLeuArgHisValAspPheTyrIleMet 219

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        220 ProValMetAsnValAspGlyTyrAspTyrThrTrpLysLysAsnArgMetTrpArgLys 239

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        240 AsnArgSerValHisMetAsnAsnArgCysValGlyThrAspLeuAsnArgAsnPheAla 259

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||:::|||||||||||||||||||||||||||||||||||||||
Db        260 SerLysHisTrpCysGluLysGlyAlaSerSerPheSerCysSerGluThrTyrCysGly 279

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        280 LeuTyrProGluSerGluProGluValLysAlaValAlaAspPheLeuArgArgAsnIle 299
```

```
Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              |||   |||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        300 AsnHisIleLysAlaTyrIleSerMetHisSerTyrSerGlnGlnIleLeuProTyr 319

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 SerTyrAsnArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 339

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||:::      :::|||||||||||||||||||||||   |||||||||
Db        340 ValArgAlaIleGluSerIleAsnLysAsnThrArgTyrThrHisGlySerGlySerGlu 359

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              :::|||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db        360 SerLeuTyrLeuAlaProGlyGlySerAspAspTrpIleTyrAspLeuGlyIleLysTyr 379

Qy       1030 TCGTTTACATC---------------------------------------------------- 1040
              |||||||||
Db        380 SerPheThrIleGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluArgPhe 399

Qy       1041 ---AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1079
                 ||||||||||||   ||||||   |||||||||||||||
Db        400 IleLysProThrCysAlaGluAlaLeuAlaAlaValSerLys 413

RESULT 7
Q9NTI8
ID   Q9NTI8      PRELIMINARY;      PRT;   198 AA.
AC   Q9NTI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN   CPB2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tromans A.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137141; CAB92622.1; -.
DR   HSSP; P00730; 1CPX.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
KW   Carboxypeptidase.
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  22460 MW;  32F005305621C2A5 CRC64;

Alignment Scores:
Pred. No.:              9.22e-95       Length:        198
Score:                  1029.00        Matches:       196
Percent Similarity:     99.49%         Conservative:  0
Best Local Similarity:  99.49%         Mismatches:    1
Query Match:            53.26%         Indels:        0
DB:                     4              Gaps:          0

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9NTI8 (1-198)

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        161 LeuLysValSerGlyLysGluGlnAlaAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197

RESULT 8
Q9PUF2
ID   Q9PUF2      PRELIMINARY;      PRT;   416 AA.
AC   Q9PUF2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase homolog.
OS   Bothrops jararaca (Jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
OX   NCBI_TaxID=8724;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT   "Screening of Bothrops jararaca pancreas cDNA library.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF190274; AAF01344.1; -.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ   SEQUENCE   416 AA;  47723 MW;  4F999854DD72B7A7 CRC64;

Alignment Scores:
Pred. No.:              3.71e-62       Length:        416
Score:                  706.00         Matches:       149
Percent Similarity:     53.77%         Conservative:  65
Best Local Similarity:  37.44%         Mismatches:    124
Query Match:            36.54%         Indels:        61
DB:                     13             Gaps:          6

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9PUF2 (1-416)

Qy         43 TTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGA 102
```

```
Db         14 PheAlaGluThrThrValHisArgPheAspGlyGluLysValTyrArgValThrProArg 33

Qy        103 ACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGG 162
Db         34 AsnGluAspGluValTyrPheLeuAsnTyrLeuAlaAsnIleValGlnValAspPheTrp 53

Qy        163 CAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCA 222
Db         54 ArgProAspSerValGluLeuValLysAlaGluMetThrValAspPheArgIleGluAla 73

Qy        223 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 282
Db         74 AspArgCysSerGluValGluSerIleLeuGlnGlnSerGlyLeuAsnTyrGluIleLeu 93

Qy        283 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 342
Db         94 IleAspAsnLeuGlnAlaValLeuAspArgGlnLeuAspAsnHisAla---------Arg 110

Qy        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402
Db        111 ThrAlaGlyTyrAsnTyrGluLysTyrAsnSerTrpGluLysIleAspAlaTrpThrAla 130

Qy        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462
Db        131 AspIleAlaAsnGluAsnProSerLeuValSerArgLeuGlnIleGlyThrThrPheGlu 150

Qy        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 522
Db        151 GlyArgProMetProLeuLeuLysVal---GlyLysProGlyValAsnLysLysAlaIle 169

Qy        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582
Db        170 PheIleAspCysGlyPheHisAlaArgGluTrpIleSerProAlaPheCysGlnTrpPhe 189

Qy        583 ATA---------------------------------------------------------- 585
Db        190 ValArgGluAlaValArgThrTyrGlyLysGluThrIleMetThrGlnLeuLeuAsnLys 209

Qy        586 ----------------------------------------------------GGCCAT 591
Db        210 LeuAspPheTyrIleLeuProValLeuAsnIleAspGlyTyrValTyrSerTrpLysGln 229

Qy        592 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 651
Db        230 SerArgMetTrpArgLysThrArgSerValAsnAlaGlySerThrCysIleGlyThrAsp 249

Qy        652 CTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 711
Db        250 ProAsnArgAsnPhe---AspAlaAlaTrpCysSerValGlyAlaSerArgAsnProCys 268

Qy        712 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 771
Db        269 SerGluThrTyrCysGlySerLysProGluSerGluLysGluThrLysAlaLeuAlaAsp 288

Qy        772 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 831
Db        289 PheIleArgArgAsnArgSerIleIleGlnAlaTyrLeuThrIleHisSerTyrSerGln 308

Qy        832 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 891
Db        309 MetLeuLeuTyrProTyrSerTyrThrTyrAspLeuThrSerAsnAsnLysLysLeuAsn 328

Qy        892 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACA 951
Db        329 SerIleAlaLysGluAlaIleArgGluLeu---LysValLeuPheGlyThrGluTyrThr 347

Qy        952 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1011
Db        348 TyrGlyProGlyAlaAlaThrIleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyr 367

Qy       1012 GATTTGGGCATCAAATATTCGTTTACATC---------------------------- 1040
```

```
Db        368 AspGlnGlyIleLysTyrAlaPheThrPheGluLeuArgAspLysGlyArgTyrGlyPhe 387

Qy       1041 --------------------AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
Db        388 AlaLeuProGluSerGlnIleLysProThrCysGluGluThrMetIleAlaVal 405

RESULT 9
Q8AXN3
ID   Q8AXN3      PRELIMINARY;      PRT;   408 AA.
AC   Q8AXN3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B (Fragment).
GN   CPB.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Srivastava A.S., Kurokawa T., Suzuki T.;
RT   "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT   A2 and B from the Japanese flounder, Paralichthys olivaceus.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB099302; BAC53789.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   408 AA;  45904 MW;  809CAACD9D2DA34B CRC64;

Alignment Scores:
Pred. No.:              2.14e-60        Length:         408
Score:                  688.50          Matches:        149
Percent Similarity:     53.85%          Conservative:   61
Best Local Similarity:  38.21%          Mismatches:     121
Query Match:            35.64%          Indels:         60
DB:                     13              Gaps:           6

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8AXN3 (1-408)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db         12 PheGluGlyGluLysValPheArgLeuLysProValPheAspGluHisValThrLeuIle 31

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
Db         32 ArgAspLeuAlaAsnSerIleGluValAspPheTrpArgProGluSerSerGlnLeuVal 51

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
Db         52 ThrIleAspIleAspValAspIleArgValProAlaIleTyrLeuAspMetValTyrThr 71

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
Db         72 LeuLeuGlnGlnSerAspMetGluHisGluValLeuIleGluAspLeuGlnSerAlaVal 91

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
Db         92 AspAlaGluAlaAsp---------LeuLysProSerProArgAlaHisSerTyrThrLys 108

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
Db        109 TyrAsnAsnLeuAspLysValGlnSerTrpIleAlaSerIleSerSerSerAsnProAsp 128

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
Db        129 LeuIleSerLysGlnValIleGlyAsnThrTyrGluGlyArgProMetThrValLeuLys 148

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
Db        149 LeuGlyLysLysSerSerSerThrLysProAlaIlePheMetAspCysGlyIleHisAla 168
```

```
Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------- 585
Db        169 ArgGluTrpIleSerProAlaPheCysGlnTrpPheValLysGluAlaLeuSerThrTyr 188

Qy        585 ------------------------------------------------------------ 585

Db        189 GlySerAspSerGlnMetThrSerLeuLeuAsnGlnMetAspValPheValLeuProVal 208

Qy        586 ------------GGCCAT-----------------AATCGAATGTGGAGAAAGAACCGT 615
                          |||:::                 ||||||||||||||||||   |||
Db        209 PheAsnIleAspGlyTyrAspPheThrHisLysSerAsnArgMetTrpArgLysThrArg 228

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
              |||      :::   :::   |||||||||   |||   |||   ||||||
Db        229 SerArgLysSerGlySerSerCysIleGlyAlaAspProAsnArgAsnPhe---AspAla 247

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
                |||||      |||||||||   :::   ||||||:::|||:::||||||
Db        248 GlyTrpCysSerMetGlyAlaSerAspAsnProCysSerAspThrPheCysGlyTyrThr 267

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
              ||||||||||||   |||||||||   ||||||   |||:::|||||||||   :::
Db        268 ProGluSerGluIleGluValLysAsnValAlaAspPheIleArgArgAsnArgSerIle 287

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
              ||||||||||||:::::::::|||||||||||||||   ::::::||||||||||||||
Db        288 IleLysAlaTyrLeuThrValHisSerTyrSerGlnLeuLeuLeuProTyrSerTyr 307

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
              |||          :::   ||||||   ||||||       ||||||       |||
Db        308 ThrTyrGlnLeuAlaAlaAspHisSerGluLeuMetLysValAlaGluGlyAlaSerAla 327

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
              |||:::   ::::::          |||   ||||||   |||   |||:::   |||:::
Db        328 AlaLeu---ArgSerLeuTyrGlyThrThrTyrThrSerGlyProGlyAlaThrThrIle 346

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
              |||   |||   ||||||   |||||||||   ||||||||||:::||||||||:::
Db        347 TyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspLeuGlyValLysTyrSerTyr 366

Qy       1036 ACATC-------------------------------------------------AAA 1043
              |||                                                    |||
Db        367 ThrPheGluLeuArgAspThrGlyArgTyrGlyPheLeuLeuProGluSerGlnIleLys 386

Qy       1044 CCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              |||||||||   |||         ||||||
Db        387 ProThrCysGluGluThrMetLeuAlaVal 396
```

```
RESULT 10
Q9XSP3
ID   Q9XSP3      PRELIMINARY;      PRT;   416 AA.
AC   Q9XSP3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Procarboxypeptidase B precursor (EC 3.4.17.2).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=99321926; PubMed=10391940;
RA   Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA   Hershberger C., Aviles F.;
RT   "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT   Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL   J. Biol. Chem. 274:19925-19933(1999).
DR   EMBL; AJ133775; CAB46991.1; -.
DR   HSSP; P09955; 1PBA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Hydrolase; Signal.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16    416       PROCARBOXYPEPTIDASE B.
SQ   SEQUENCE   416 AA;  47381 MW;  5FA7FCED7B45AB6A CRC64;
```

```
Alignment Scores:
Pred. No.:              4.1e-56        Length:          416
Score:                  646.00         Matches:         140
Percent Similarity:     53.40%         Conservative:    64
Best Local Similarity:  36.65%         Mismatches:      134
Query Match:            33.44%         Indels:          44
DB:                     6              Gaps:            7

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q9XSP3 (1-416)

Qy          7 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 66
              :::   :::      :::||||||   :::   |||              |||
Db          1 MetLeuAlaPheLeuIleLeuValThrValThrLeuAlaSerAlaHisHisSerGlyGlu 20

Qy         67 ---TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
                 |||:::      :::|||                                  :::   :::
Db         21 HisPheGluGlyGluLysValPheArgValAsnValGluAspGluAsnAspIleSerLeu 40

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 183
              |||      |||   :::|||   :::|||      |||:::|||   :::
Db         41 LeuHisGluLeuAlaSerThrArgGlnIleAspPheTrpLysProAspSerValThrGln 60

Qy        184 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 243
              |||                |||   |||   |||   |||   |||:::      |||:::
Db         61 IleLysProHisSerThrValAspPheArgValLysAlaGluAspIleLeuAlaValGlu 80

Qy        244 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 303
                  |||      :::   :::         ||||||:::   ::::::      :::
Db         81 AspPheLeuGluGlnAsnGluLeuGlnTyrGluValLeuIleAsnAsnLeuArgSerVal 100

Qy        304 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ::::::   |||      :::                |||   :::         ||||||
Db        101 LeuGluAlaGlnPheAspSerArgVal---------ArgThrThrGlyHisSerTyrGlu 117

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              :::|||::::::          |||   :::|||   :::   :::|||      :::|||
Db        118 LysTyrAsnAsnTrpGluThrIleGluAlaTrpThrLysGlnValThrSerGluAsnPro 137

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              |||::::::::::::       ||||||::::::|||             :::|||:::|||
Db        138 AspLeuIleSerArgThrAlaIleGlyThrThrPheLeuGlyAsnAsnIleTyrLeuLeu 157

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              ||||||   ||||||          |||   ||||||::::::|||||||||   |||
Db        158 LysVal---GlyLysProGlyProAsnLysProAlaIlePheMetAspCysGlyPheHis 176

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 585
              |||||||||||||||||   |||||||||   ||||||:::
Db        177 AlaArgGluTrpIleSerHisAlaPheCysGlnTrpPheValArgGluAlaValLeuThr 196

Qy        585 ------------------------------------------------------------ 585

Db        197 TyrGlyTyrGluSerHisMetThrGluPheLeuAsnLysLeuAspPheTyrValLeuPro 216

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 ||||||||||||||||||
Db        217 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrLysAsnArgMetTrpArgLysThr 236

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||      |||         |||||||||||||||   |||   ||||||
Db        237 ArgSerThrAsnAlaGlyThrThrCysIleGlyThrAspProAsnArgAsnPhe---Asp 255
```

```
Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
Db        256 AlaGlyTrpCysThrThrGlyAlaSerThrAspProCysAspGluThrTyrCysGlySer 275

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
Db        276 AlaAlaGluSerGluLysGluThrLysAlaLeuAlaAspPheIleArgAsnAsnLeuSer 295

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
Db        296 SerIleLysAlaTyrLeuThrIleHisSerTyrSerGlnMetIleLeuTyrProTyrSer 315

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
Db        316 TyrAspTyrLysLeuProGluAsnAsnAlaGluLeuAsnAsnLeuAlaLysAlaAlaVal 335

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
Db        336 LysGluLeu---AlaThrLeuTyrGlyThrLysTyrThrTyrGlyProGlyAlaThrThr 354

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
Db        355 IleTyrProAlaAlaGlyGlySerAspAspTrpAlaTyrAspGlnGlyIleLysTyrSer 374

Qy       1033 TTTACA 1038
Db        375 PheThr 376

RESULT 11
P97597
ID   P97597      PRELIMINARY;      PRT;   412 AA.
AC   P97597;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (Fragment).
GN   R-CPA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague Dawley;
RX   MEDLINE=97149430; PubMed=8996238;
RA   Lutzelschwab C., Pejler G., Aveskogh M., Hellman L.;
RT   "Secretory granule proteases in rat mast cells. Cloning of 10
RT   different serine proteases and a carboxypeptidase A from various rat
RT   mast cell populations.";
RL   J. Exp. Med. 185:13-29(1997).
DR   EMBL; U67914; AAB48267.1; -.
DR   HSSP; P09955; 1NSA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Signal.
FT   NON_TER       1     1
FT   SIGNAL       <1     10       POTENTIAL.
FT   CHAIN        11     412      MAST CELL CARBOXYPEPTIDASE A.
SQ   SEQUENCE   412 AA;  47944 MW;  DAB59555FC49137D CRC64;

Alignment Scores:
Pred. No.:              1.49e-54       Length:          412
Score:                  630.50         Matches:         141
Percent Similarity:     50.77%         Conservative:    57
Best Local Similarity:  36.15%         Mismatches:      133
Query Match:            32.63%         Indels:          60
DB:                     11             Gaps:            7
```

```
US-09-980-881A-1_COPY_18_1097 (1-1080) x P97597 (1-412)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
Db         16 PheAspArgGluLysValPheArgValLysLeuGlnAspGluLysGlnAlaSerIleLeu 35

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
Db         36 LysAsnLeuThrGlnThrIleGluLeuAspPheTrpTyrProAspAlaIleHisAspIle 55

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246
Db         56 AlaValAsnMetThrValAspPheArgValThrGluLysGluSerGlnThrIleGlnSer 75

Qy        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306
Db         76 ThrLeuGluGlnHisLysMetAspTyrGluIleLeuIleAsnAspLeuGlnGluGluIle 95

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
Db         96 AspLysGlnPhe------AspValLysGluGluIleAlaGlyArgHisSerTyrAlaLys 113

Qy        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426
Db        114 TyrAsnAspTrpAsnLysIleValSerTrpThrGluLysMetValGluLysHisProGlu 133

Qy        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486
Db        134 MetValSerArgIleLysIleGlySerThrValGluAspAsnProLeuTyrValLeuLys 153

Qy        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546
Db        154 Ile---GlyArgLysAspGlyGluArgLysAlaIlePheMetAspCysGlyIleHisAla 172

Qy        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------- 585
Db        173 ArgGluTrpValSerProAlaPheCysGlnTrpPheValTyrGlnAlaAlaLysSerTyr 192

Qy        586 GGCCATAAT---------- 594
Db        193 GlyLysAsnLysIleMetThrLysLeuLeuAspArgMetAsnPheTyrValLeuProVal 212

Qy        595 ----------CGAATGTGGAGAAAGAACCGT 615
Db        213 PheAsnValAspGlyTyrIleTrpSerTrpThrLysAspArgMetTrpArgLysAsnArg 232

Qy        616 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAA 675
Db        233 SerLysAsnProAsnSerThrCysIleGlyThrAspLeuAsnArgAsnPhe---AspVal 251

Qy        676 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 735
Db        252 SerTrpAspSerSerProAsnThrAspAsnProCysLeuSerValTyrArgGlyProAla 271

Qy        736 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 795
Db        272 ProGluSerGluLysGluThrLysAlaValThrAsnPheIleArgSerHisLeuAsnSer 291

Qy        796 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 855
Db        292 IleLysAlaTyrIleThrPheHisSerTyrSerGlnMetLeuLeuPheProTyrGlyTyr 311

Qy        856 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 915
Db        312 ThrIleLysLeuProProAsnHisGlnAspLeuLeuLysValAlaArgIleAlaThrAsp 331

Qy        916 GCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 975
Db        332 ValLeu---SerSerArgTyrGluThrArgTyrIleTyrGlyProIleAlaSerThrIle 350

Qy        976 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1035
```

```
Db        351 TyrLysThrSerGlySerSerLeuAspTrpAlaTyrAspLeuGlyIleLysHisThrPhe 370

Qy       1036 AC-------------------------------------------------------ATCAAA 1043
                                                                         ||||||
Db        371 AlaPheGluLeuArgAspLysGlyLysSerGlyPheLeuLeuProGluSerArgIleLys 390

Qy       1044 CCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
              |||||||||:::|||         :::|||
Db        391 ProThrCysLysGluThrMetLeuSerVal 400

RESULT 12
Q8N4T0
ID   Q8N4T0      PRELIMINARY;      PRT;   437 AA.
AC   Q8N4T0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033684; AAH33684.1; -.
DR   Genew; HGNC:17245; CPA6.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   437 AA;  51024 MW;  9970680D429A0A98 CRC64;

Alignment Scores:
Pred. No.:              1.24e-52      Length:          437
Score:                  611.50        Matches:         142
Percent Similarity:     49.76%        Conservative:    67
Best Local Similarity:  33.81%        Mismatches:      140
Query Match:            31.65%        Indels:          72
DB:                     4             Gaps:            11

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8N4T0 (1-437)

Qy           1 ATGAAGCTTTGC------AGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAG 54
                 :::   ||||||         |||   :::|||   |||
Db          15 LeuProLeuCysTrpLeuPheLeuLysIleLeuGlnPro-----------GlyHisSer 30

Qy          55 CATGTCTTC------GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCT 108
               |||::::::         :::         :::|||:::      :::|||:::|||
Db          31 HisLeuTyrAsnAsnArgTyrAlaGlyAspLysValIleArgPheIleProLysThrGlu 50

Qy         109 AGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCG 168
                  :::        |||:::   ::::::        ::::::   ||||||||||||
Db          51 GluGluAlaTyrAlaLeuLysLysIleSerTyrGlnLeuLysValAspLeuTrpGlnPro 70

Qy         169 GTAACAGCTGACCTTATTGTGAAG------AAAAAACAAGTCCATTTTTTTGTAAATGCA 222
                  :::         :::   :::            ||||||         |||
Db          71 SerSerIleSerTyrValSerGluGlyThrValThrAspValHisIleProGlnAsnGly 90

Qy         223 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 282
               |||           :::   |||   |||     :::   |||         ||||||
Db          91 SerArgAla------LeuLeuAlaPheLeuGlnGluAlaAsnIleGlnTyrLysValLeu 108

Qy         283 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 342
                  :::   |||::::::      :::::::::      |||      |||         |||
Db         109 IleGluAspLeuGlnLysThrLeuGluLysGlySerSerLeuHisThrGlnArgAsnArg 128

Qy         343 GCCTCCGCATCG------TACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGG 396
                  |||   |||         ||||||   ||||||||||||   ||||||   :::|||
Db         129 ArgSerLeuSerGlyTyrAsnTyrGluValTyrHisSerLeuGluGluIleGlnAsnTrp 148

Qy         397 ATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCA 456
                  :::      :::   :::   |||      ::::::          ||||||   |||
Db         149 MetHisHisLeuAsnLysThrHisSerGlyLeuIleHisMetPheSerIleGlyArgSer 168

Qy         457 TTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAAT 516
                  :::|||         |||::::::||||||:::   |||:::            |||
Db         169 TyrGluGlyArgCysLeuPheIleLeuLysLeu---GlyArgArgSerArgLeuLysArg 187

Qy         517 GCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTG 576
                  |||:::||||||||||||||||||||||||||||||||||   ||||||||||||
Db         188 AlaValTrpIleAspCysGlyIleHisAlaArgGluTrpIleGlyProAlaPheCysGln 207

Qy         577 TGGTTCATA---------------------------------------------------- 585
                  ||||||:::
Db         208 TrpPheValLysGluAlaLeuLeuThrTyrLysSerAspProAlaMetArgLysMetLeu 227

Qy         585 ---------------------------------------------------------- 585

Db         228 AsnHisLeuTyrPheTyrIleMetProValPheAsnValAspGlyTyrHisPheSerTrp 247

Qy         586 GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGA 645
                  ::::::|||   ||||||||   ||||||      :::          |||   |||
Db         248 ThrAsnAspArgPheTrpArgLysThrArgSerArgAsnSerArgPheArgCysArgGly 267

Qy         646 ACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCC 705
                  |||   |||   |||:::      |||   ||||||:::||||||||||||
Db         268 ValAspAlaAsnArgAsnTrpLysValLys---TrpCysAspGluGlyAlaSerMetHis 286

Qy         706 TCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTG 765
                  |||   :::|||||||||||||   :::|||||||||||||||||||||||||||||
Db         287 ProCysAspAspThrTyrCysGlyProPheProGluSerGluProGluValLysAlaVal 306

Qy         766 GCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATAC 825
                  |||:::|||||||||::::::          |||:::||||||:::|||   |||:::|||
Db         307 AlaAsnPheLeuArgLysHisArgLysHisIleArgAlaTyrLeuSerPheHisAlaTyr 326

Qy         826 TCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAA 885
                  :::|||   :::::::::||||||||||||      :::          :::
Db         327 AlaGlnMetLeuLeuTyrProTyrSerTyrLysTyrAlaThrIleProAsnPheArgCys 346

Qy         886 CTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGG 945
                  :::          |||   :::||||||   |||:::   ::::::            |||
Db         347 ValGluSerAlaAlaTyrLysAlaValAsnAlaLeu---GlnSerValTyrGlyValArg 365

Qy         946 TATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGG 1005
                  |||   :::|||      |||   |||||||||::::::   |||         ||||||
Db         366 TyrArgTyrGlyProAlaSerThrThrLeuTyrValSerSerGlySerSerMetAspTrp 385

Qy        1006 ATCTATGATTTGGGCATCAAATATTCGTTTAC--------------------------- 1037
                  |||      ||||||   |||:::|||
Db         386 AlaTyrLysAsnGlyIleProTyrAlaPheAlaPheGluLeuArgAspThrGlyTyrPhe 405

Qy        1038 ----------------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                                       ||||||||||||||   |||         ||||||
Db         406 GlyPheLeuLeuProGluMetLeuIleLysProThrCysThrGluThrMetLeuAlaVal 425

RESULT 13
Q8AVS2
ID   Q8AVS2      PRELIMINARY;      PRT;   434 AA.
AC   Q8AVS2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041315; AAH41315.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   434 AA;  50581 MW;  B42BD77843910591 CRC64;

Alignment Scores:
Pred. No.:              1.75e-52       Length:        434
Score:                  610.00         Matches:       145
Percent Similarity:     49.41%         Conservative:  63
Best Local Similarity:  34.44%         Mismatches:    134
Query Match:            31.57%         Indels:        80
DB:                     13             Gaps:          11

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8AVS2 (1-434)

Qy         22 GTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAA 81
              ||||||:::   :::::|||   |||          |||      :::||||||
Db         11 ValLeuLeuPheLeuLeuLeuSerCys--------------AlaLeuSerAlaGlyGln 25

Qy         82 ----------------------------------GTTCTAGCTGCTCTTCCTAGAACC 105
                                                 |||:::        |||:::
Db         26 CysHisLeuTyrAsnAsnArgTyrSerTrpAspLysValIleArgValPheProLysAsn 45

Qy        106 TCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAG 165
                ||:::        |||:::::::::          ::::::   |||||||||
Db         46 GluArgGluAlaAsnGluLeuLysAspMetTyrGlnGlnLeuGlnValAspLeuTrpGln 65

Qy        166 CCGGTAACAGCTGACCTTATTGTGAAGAAA------AAACAAGTCCATTTTTTTGTAAAT 219
              |||   :::         |||   |||              ||||||         |||
Db         66 ProSerSerIleSerHisIleGlyLysAspThrValThrAspValHisThrSerGlyAsn 85

Qy        220 GCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTC 279
              :::|||:::|||      :::     :::|||     :::   |||       :::
Db         86 SerSerGluVal------LeuLeuThrTyrLeuThrLysAlaLysIleGlnHisLysIle 103

Qy        280 TTGCTGGCAGACGTGGAAGATCTTATTCAA---CAGCAGATTTCCAACGACACAGTCAGC 336
              |||:::   :::|||:::   :::::::::   ||||||
Db        104 LeuValAsnAsnValGlnSerMetLeuGluAlaGlnGlnAlaPheArgProArgArgLys 123

Qy        337 CCCCGAGCCTCCGCATCGTAC---TATGAACAGTATCACTCACTAAATGAAATCTATTCT 393
              |||:::   :::   |||   |||   :::||||||   |||:::||||||   |||
Db        124 ArgArgSerLeuSerArgTyrAsnTyrAsnGluTyrHisProLeuHisGluIleGluSer 143

Qy        394 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 453
              |||:::   ::::::   :::   |||   |||:::::::::         ||||||
Db        144 TrpMetPheTyrMetAsnLysThrHisHisAspLeuValSerLeuPheThrIleGlyLys 163

Qy        454 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAA 513
              |||:::|||         |||||||||||||||:::            ::::::   |||
Db        164 SerTyrGluGlyArgSerLeuTyrValLeuLysLeuGlyLysAspThrLysSerTyrLys 183

Qy        514 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGC 573
              |||||||||||||||||||:::|||||||||||||||||   ||||||||||||
Db        184 LysAlaIleTrpIleAspCysGlyMetHisAlaArgGluTrpIleGlyProAlaPheCys 203

Qy        574 TTGTGGTTCATA---------------------------------------------- 585
              ||||||:::
Db        204 GlnTrpPheValLysGluAlaIleAsnSerTyrAsnThrAspProAlaMetLysLysIle 223

Qy        586 ------------------------------------------GGCCAT--------- 591
                                                        |||:::
Db        224 LeuAsnLeuLeuTyrIleTyrValMetProValPheAsnValAspGlyTyrHisTyrSer 243

Qy        592 ---------AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 642
                      :::|||   |||||||||   ||||||                  |||
Db        244 TrpHisSerAspArgPheTrpArgLysThrArgSerLysAsnThrArgTyrGlnCysTyr 263

Qy        643 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 702
              |||   |||   |||   |||:::         ||||||   :::|||||||||||
Db        264 GlyValAspAlaAsnArgAsnTrp---LysValHisTrpSerAspGluGlyAlaSerLeu 282

Qy        703 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 762
              :::   |||       ||||||||||||   |||   |||||||||||||||||||||
Db        283 AsnProCysAspAsnThrTyrCysGlyProTyrAlaGluSerGluProGluValLysAla 302

Qy        763 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 822
              ||||||   ||||||   :::                 ::::::||||||:::|||   |||:::
Db        303 ValAlaGlnPheLeuTyrLysGlnArgLysHisValArgAlaTyrMetSerPheHisAla 322

Qy        823 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 882
              |||:::|||   :::::::::||||||||||||                   :::
Db        323 TyrAlaGlnMetLeuLeuTyrProTyrSerTyrGlnTyrGlyAlaIleProAsnPheGly 342

Qy        883 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 942
                 :::         |||       ||||||   ||||||   :::::::::
Db        343 CysValGluSerAlaAlaHisAsnAlaValLeuAlaIle---ArgSerAlaTyrGlyIle 361

Qy        943 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1002
              ||||||   ||||||       |||   ||||||||||||      |||         |||
Db        362 ArgTyrArgHisGlyProAlaSerSerThrLeuTyrLeuThrSerGlySerSerMetAsp 381

Qy       1003 TGGATCTATGATTTGGGCATCAAATATTCGTTTAC------------------------ 1037
              |||   |||:::   ||||||   ||||||:::
Db        382 TrpAlaTyrAsnAsnGlyIleProTyrSerTyrAlaPheGluLeuArgAspThrGlyTyr 401

Qy       1038 --------------------------ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 1070
                                         ||||||||||||||   |||         |||
Db        402 TyrGlyPheLeuLeuProGluGlyLeuIleLysProThrCysValGluThrMetLeuAla 421

Qy       1071 GTC 1073
              |||
Db        422 Val 422

RESULT 14
Q8UUK1
ID   Q8UUK1      PRELIMINARY;      PRT;   419 AA.
AC   Q8UUK1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Preprocarboxypeptidase A precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=domesticus; TISSUE=Pancreas;
RA   Hasegawa S., Honda K., Hikami Y.;
RT   "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
RT   precursor.";
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X64539; CAA45837.1; -.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Signal; Carboxypeptidase.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN      113    419       CARBOXYPEPTIDASE A.
SQ   SEQUENCE   419 AA;  47006 MW;  299C334274BCBC9C CRC64;

Alignment Scores:
Pred. No.:              1.45e-49       Length:        419
Score:                  581.00         Matches:       133
Percent Similarity:     52.52%         Conservative:  65
Best Local Similarity:  35.28%         Mismatches:    107
```

```
Query Match:          30.07%      Indels:        72
DB:                   13          Gaps:          9

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8UUK1 (1-419)

Qy         67 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126
              |||         |||||||||      :::|||   :::      :::::::|||   :::
Db         21 PheValGlyHisGlnValLeuArgIleValProSerSerAspAlaGluLeuGlnLysVal 40

Qy        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186
              |||   |||
Db         41 GlnGluLeuGlnGlu---------------------------------------------- 45

Qy        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAAT--------- 237
              ::::::      |||:::   |||::::::::::         :::   |||
Db         46 LeuGluHisLeuGlnLeuAspPheTrpLeuSerProArgGlyLeuGlyAsnProValAsp 65

Qy        238 ----------------------------GTGAAAGCCCATTTAAATGTGAGCGGAATT 267
                                          ||||||||||||||      :::|||:::
Db         66 IleArgValProPheProSerLeuGlnProValLysAlaHisLeuGluAlaAsnGlyVal 85

Qy        268 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT----------CAACAGCAG 315
              |||   |||::::::::::   ||||||:::  |||:::           |||
Db         86 ProTyrSerIleMetIleGluAspValGlnAlaLeuValAspLeuGluGlnMetGlnMet 105

Qy        316 ATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAACAGTATCAC 372
              :::                  |||   ::::::::::::::   |||      ||||||
Db        106 LeuArgArgArgArgPheValProLeuSerThrSerThrPheAspTyrThrSerTyrHis 125

Qy        373 TCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTT 432
              :::|||   |||||||||:::::::::::  :::         :::|||:::::::::
Db        126 ThrLeuGluGluIleTyrAlaPheMetAspLeuLeuValAlaGluAsnProAsnLeuVal 145

Qy        433 ACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCT 492
              :::|||:::   ||||||  :::   |||      |||:::||||||||||||  |||
Db        146 SerLysLeuGluIleGlyArgThrThrGluAsnArgProIleTyrValLeuLysPheSer 165

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
                 |||      |||   :::   ||||||||||||||   |||||||||:::||||||
Db        166 ---LysGlyGlyThrAsnArgProAlaIleTrpIleAspThrGlyIleHisSerArgGlu 184

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA-------------------------- 585
              |||::::::   |||      :::||||||
Db        185 TrpValThrGlnAlaSerGlyIleTrpPheAlaLysLysIleValGlnGluGlnAspGlu 204

Qy        585 ------------------------------------------------------------ 585

Db        205 GlyLeuAlaAsnIleLeuAspGlnMetAspIlePheLeuGluIleValThrAsnProAsp 224

Qy        586 ------------------GGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCG 627
                                |||||||||||||||||   ||||||      :::
Db        225 GlyPheAlaPheThrHisThrSerAsnArgMetTrpArgLysThrArgSerLysArgSer 244

Qy        628 AACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAG 687
              :::   |||:::|||   |||   |||   |||:::          :::
Db        245 GlySerLeuCysValGlyValAspProAsnArgAsnTrp---AspAlaGlyPheGlyGly 263

Qy        688 GAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAA 747
              ||||||||||:::   ||||||||||||||   |||   |||      ||||||
Db        264 SerGlyAlaSerSerAsnProCysSerGluThrTyrHisGlyProTyrAlaAsnSerGlu 283

Qy        748 CCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATAC 807
              |||||||||||||:::      |||::::::   :::   |||   |||||||||:::
Db        284 ProGluValLysAlaIleValAspPheValLysSerHisLysAsn---IleLysAlaPhe 302

Qy        808 ATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAA 867
              ||||||:::|||||||||||||||   :::::::::||||||   ||||||   :::
Db        303 IleSerIleHisSerTyrSerGlnLeuLeuLeuTyrProTyrGlyTyrThrThrThrAla 322

Qy        868 AGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAA 927
                  |||   |||||||||      ||||||   |||||||||   |||:::
Db        323 ValProAspLysGluGluLeuHisGlnValAlaLysGluAlaValAlaAlaLeu---Ser 341

Qy        928 ACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCT 987
              :::         |||   |||   :::|||            |||:::|||   |||
Db        342 SerLeuTyrGlyThrAsnTyrLysTyrGlySerIleIleThrThrIleTyrGlnAlaSer 361

Qy        988 GGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTACA 1038
              ||||||      ||||||   |||:::   ||||||||||||||||||||
Db        362 GlyGlyThrIleAspTrpThrTyrAsnGlnGlyIleLysTyrSerPheThr 378

RESULT 15
Q8IVL8
ID   Q8IVL8      PRELIMINARY;      PRT;   374 AA.
AC   Q8IVL8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zn-carboxypeptidase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Obaya A.J., Lopez-Otin C.;
RT   "A new Zn-carboxypeptidase highly expressed in ovary.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ422118; CAD19478.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   374 AA;  42529 MW;  404C373BB841AAD2 CRC64;

Alignment Scores:
Pred. No.:          6.49e-48      Length:        374
Score:              564.50        Matches:       115
Percent Similarity: 53.70%        Conservative:  52
Best Local Similarity: 36.98%     Mismatches:    87
Query Match:        29.22%        Indels:        58
DB:                 4             Gaps:          6

US-09-980-881A-1_COPY_18_1097 (1-1080) x Q8IVL8 (1-374)

Qy        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTAC---TATGAA 363
              :::|||:::|||   :::   :::|||||||||   :::      :::|||   |||
Db         29 ArgGlnGluIleValAspLysSerValSerProTrpSerLeuGluThrTyrSerTyrAsn 48

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              ||||||   :::   |||||||||   |||:::      |||:::|||::::::
Db         49 IleTyrHisProMetGlyGluIleTyrGluTrpMetArgGluIleSerGluLysTyrLys 68

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              :::::::::||:::      :::|||   ::::::|||   :::|||:::|||   |||
Db         69 GluValValThrGlnHisPheLeuGlyValThrTyrGluThrHisProMetTyrTyrLeu 88

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 543
              |||:::|||                  |||      ||||||:::||||||||||||||||
Db         89 LysIleSerGlnProSerGlyAsnProLysLysIleIleTrpMetAspCysGlyIleHis 108

Qy        544 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATA---------------- 585
              ||||||||||||||:::||||||||||||   ||||||:::
Db        109 AlaArgGluTrpIleAlaProAlaPheCysGlnTrpPheValLysGluIleLeuGlnAsn 128

Qy        585 ------------------------------------------------------------ 585

Db        129 HisLysAspAsnSerSerIleArgLysLeuLeuArgAsnLeuAspPheTyrValLeuPro 148

Qy        586 ---------------GGCCAT-----------------AATCGAATGTGGAGAAAGAAC 612
                            |||:::                 :::|||:::|||||||||:::
Db        149 ValLeuAsnIleAspGlyTyrIleTyrThrTrpThrThrAspArgLeuTrpArgLysSer 168

Qy        613 CGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCC 672
              ||||||   :::   |||      |||   ||||||||||||||   ||||||   :::
Db        169 ArgSerProHisAsnAsnGlyThrCysPheGlyThrAspLeuAsnArgAsnPhe---Asn 187

Qy        673 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 732
```

```
                    ||||||      |||||||||   :::         :::|||:::||||||
Db        188 AlaSerTrpCysSerIleGlyAlaSerArgAsnCysGlnAspGlnThrPheCysGlyThr 207

Qy        733 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAAC 792
                 |||   ||||||||||||   ||||||||||||||||||:::            :::
Db        208 GlyProValSerGluProGluThrLysAlaValAlaSerPheIleGluSerLysLysAsp 227

Qy        793 CAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCC 852
                 |||      :::::::::|||||||||||||   |||   |||:::   ||||||
Db        228 AspIleLeuCysPheLeuThrMetHisSerTyrGlyGlnLeuIleLeuThrProTyrGly 247

Qy        853 TATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTT 912
                 ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db        248 TyrThrLysAsnLysSerSerAsnHisProGluMetIleGlnValGlyGlnLysAlaAla 267

Qy        913 CGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACC 972
                 |||:::   |||           |||   |||      |||      ::::::
Db        268 AsnAlaLeu---LysAlaLysTyrGlyThrAsnTyrArgValGlySerSerAlaAspIle 286

Qy        973 TTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCG 1032
                 ||||||   :::   |||          ||||||      |||:::||||||   :::|||
Db        287 LeuTyrAlaSerSerGlySerSerArgAspTrpAlaArgAspIleGlyIleProPheSer 306

Qy       1033 TTTACATC---------------------------------------------------- 1040
                 :::|||
Db        307 TyrThrPheGluLeuArgAspSerGlyThrTyrGlyPheValLeuProGluAlaGlnIle 326

Qy       1041 AAACCCACCTGTAGAGAAGCTTTTGCCGCTGTC 1073
                 :::|||||||||   |||         ||||||
Db        327 GlnProThrCysGluGluThrMetGluAlaVal 337


Search completed: January  1, 2004, 18:58:53
Job time : 74.6425 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:58:24 ; Search time 4065.16 Seconds
(without alignments)
10868.540 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1064	98.5	1546	9	AB011969	AB011969 Homo sapi
2		1025.2	94.9	1625	6	AR232183	AR232183 Sequence
3		909	84.2	1272	6	AR086324	AR086324 Sequence
4		909	84.2	1272	6	BD106653	BD106653 Method of
5		909	84.2	1728	6	AX409750	AX409750 Sequence
6		909	84.2	1728	6	AX706330	AX706330 Sequence
7		909	84.2	1728	9	HUMPCPBX	M75106 Human prepr
8		909	84.2	1749	6	I16100	I16100 Sequence 2
9		909	84.2	1749	6	I33526	I33526 Sequence 2
10		907.4	84.0	1272	9	BT006936	BT006936 Homo sapi
11		907.4	84.0	1272	12	BT007990	BT007990 Synthetic
12		907.4	84.0	1715	9	BC007057	BC007057 Homo sapi
13		643.4	59.6	1421	10	AB021968	AB021968 Mus muscu
14		643.4	59.6	1490	10	AF164524	AF164524 Mus muscu
15		642.4	59.5	1425	10	AB042598	AB042598 Rattus no
16		638.6	59.1	1430	10	AF186188	AF186188 Mus muscu
17		204	18.9	55827	6	AR232184	AR232184 Sequence
c 18		204	18.9	58097	9	AL157758	AL157758 Human DNA
c 19		156	14.4	251806	2	AC096305	AC096305 Rattus no
20		156	14.4	258265	2	AC120472	AC120472 Rattus no
c 21		129.6	12.0	431	6	AX400792	AX400792 Sequence
22		128	11.9	180680	9	AL137141	AL137141 Human DNA
23		124	11.5	927	6	AR072911	AR072911 Sequence
24		123.8	11.5	1327	5	AF190274	AF190274 Bothrops
25		123.4	11.4	921	6	I67699	I67699 Sequence 3
26		123.4	11.4	1215	6	I67698	I67698 Sequence 1
27		123.4	11.4	1251	4	SSC133775	AJ133775 Sus scrof
c 28		123.2	11.4	451	11	G53326	G53326 SHGC-82320
29		120.8	11.2	1251	4	DOGZAP47	D78348 Canis famil
30		109	10.1	1263	6	A87530	A87530 Sequence 11
31		107.6	10.0	1622	9	HUMCARMC	M27717 Human mast
32		107.6	10.0	1622	11	G28614	G28614 human STS S
33		107.6	10.0	1674	9	BC012613	BC012613 Homo sapi
34		107.4	9.9	999	6	A51908	A51908 Sequence 72
35		107.4	9.9	999	6	AR085871	AR085871 Sequence
36		107.4	9.9	1053	6	A51904	A51904 Sequence 68
37		107.4	9.9	1053	6	AR085868	AR085868 Sequence
38		107.4	9.9	1254	12	BT008111	BT008111 Synthetic
39		107.4	9.9	1263	6	A51896	A51896 Sequence 60
40		107.4	9.9	1263	6	AR085861	AR085861 Sequence
41		107.4	9.9	1284	6	A51913	A51913 Sequence 77
42		107.4	9.9	1284	6	AR085875	AR085875 Sequence
43		107.4	9.9	1310	9	HSA224866	AJ224866 Homo sapi
44		107.4	9.9	1454	9	BC015338	BC015338 Homo sapi
45		104.2	9.6	1053	6	AR300428	AR300428 Sequence

ALIGNMENTS

RESULT 1
AB011969
LOCUS AB011969 1546 bp mRNA linear PRI 02-FEB-2000
DEFINITION Homo sapiens mRNA for carboxypeptidase B-like protein, complete cds.
ACCESSION AB011969
VERSION AB011969.1 GI:6855463
KEYWORDS carboxypeptidase B-like protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
TITLE Isolation, molecular cloning, and partial characterization of a

```
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
REFERENCE   2  (sites)
  AUTHORS   Matsumoto,A.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1546)
  AUTHORS   Matsumoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
            Medicine, Department of Radiation Biophysics & Genetics;
            Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:amat@med.kobe-u.ac.jp, Tel:81-78341-7451)
COMMENT     Sequence updated (10-Apr-1998).
FEATURES             Location/Qualifiers
     source          1. .1546
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             20. .1102
                     /note="carboxypeptidase B-like protein in human brain"
                     /codon_start=1
                     /product="carboxypeptidase B-like protein"
                     /protein_id="BAA90475.1"
                     /db_xref="GI:6855464"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQTGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDL
                     NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYS
                     QHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDD
                     WIYDLGIKYSFTSNPPVEKLLPLSLK"
     polyA_site      1546
                     /note="21 A nucleotides"
BASE COUNT      453 a    323 c    297 g    473 t
ORIGIN
```

Query Match 98.5%; Score 1064; DB 9; Length 1546;
Best Local Similarity 99.1%; Pred. No. 2.1e-290;
Matches 1070; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| |||| ||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGACTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 319

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATT 559

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 619

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 679

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 739

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 799

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 859

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 919

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              |||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||||
Db        920 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 979

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1039

Qy       1021 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1099
```

```
RESULT 2
AR232183
LOCUS       AR232183                1625.bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6455294.
ACCESSION   AR232183
VERSION     AR232183.1  GI:27274083
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1625)
  AUTHORS   Gan,W., Ketchum,K.A., Di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human protease proteins, nucleic acid molecules encoding
            human protease proteins, and uses thereof
  JOURNAL   Patent: US 6455294-A 1 24-SEP-2002;
FEATURES             Location/Qualifiers
     source          1. .1625
                     /organism="unknown"
BASE COUNT      488 a    334 c    315 g    488 t
ORIGIN
```

Query Match 94.9%; Score 1025.2; DB 6; Length 1625;
Best Local Similarity 99.2%; Pred. No. 2.1e-279;
Matches 1030; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76
```

```
Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 316

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 556

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 616

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 676

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 736

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 796

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 856

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              |||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db        917 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 976

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1036

Qy       1021 ATCAAATATTCGTTTACA 1038
              |||||||||||||||||
Db       1037 ATCAAATATTCGTTTACA 1054
```

RESULT 3
AR086324
LOCUS AR086324 1272 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from patent US 5985562.

ACCESSION AR086324
VERSION AR086324.1 GI:10013090
KEYWORDS .
SOURCE Unknown.
ORGANISM Unknown.
Unclassified.
REFERENCE 1 (bases 1 to 1272)
AUTHORS Morser,M.John. and Nagashima,M.
TITLE Method of detecting thrombotic disease risk associated with plasma carboxypeptidase B polymorphisms
JOURNAL Patent: US 5985562-A 1 16-NOV-1999;
FEATURES Location/Qualifiers
source 1. .1272
/organism="unknown"
BASE COUNT 375 a 269 c 271 g 357 t
ORIGIN

Query Match 84.2%; Score 909; DB 6; Length 1272;
Best Local Similarity 89.9%; Pred. No. 1.9e-246;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        592 ---------------------------------------------------------- 591

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        592 -----------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780
```

```
Qy       670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db       781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
Db       841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
Db       901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db       961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
Db      1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
Db      1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy      1030 TCGTTTACA 1038
Db      1141 TCGTTTACA 1149
```

```
RESULT 4
BD106653
LOCUS       BD106653                1272 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Method of detecting thrombotic disease risk.
ACCESSION   BD106653
VERSION     BD106653.1  GI:23201471
KEYWORDS    JP 2002502253-A/1.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.J. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk
  JOURNAL   Patent: JP 2002502253-A 1 22-JAN-2002;
            SCHERING AG
COMMENT     PN   JP 2002502253-A/1
            PD   22-JAN-2002
            PF   02-JUN-1998 JP 1999501465
            PR   03-JUN-1997 US    08/869057
            PI   MICHAEL JOHN MORSER,MARIKO NAGASHIMA
            PC   C12Q1/68
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES            Location/Qualifiers
     source         1. .1272
                    /organism="synthetic construct"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:32630"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN

  Query Match            84.2%;  Score 909;  DB 6;  Length 1272;
  Best Local Similarity  89.9%;  Pred. No. 1.9e-246;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        592 ------------------------------------------------------------ 591
Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1030 TCGTTTACA 1038
Db       1141 TCGTTTACA 1149
```

```
RESULT 5
AX409750
LOCUS       AX409750               1728 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2397 from Patent WO0229103.
ACCESSION   AX409750
VERSION     AX409750.1  GI:21442455
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2397 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M75106"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

  Query Match             84.2%;  Score 909;  DB 6;  Length 1728;
  Best Local Similarity   89.9%;  Pred. No. 1.9e-246;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        592 ------------------------------------------------------------ 591

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679
```

```
Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1160 TCGTTTACA 1168
```

```
RESULT 6
AX706330
LOCUS       AX706330               1728 bp    DNA     linear   PAT 04-APR-2003
DEFINITION  Sequence 1 from Patent WO03014391.
ACCESSION   AX706330
VERSION     AX706330.1  GI:29562754
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Morten,J.E. and Jackson,K.N.
  TITLE     Method
  JOURNAL   Patent: WO 03014391-A 1 20-FEB-2003;
            AstraZeneca AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

  Query Match             84.2%;  Score 909;  DB 6;  Length 1728;
  Best Local Similarity   89.9%;  Pred. No. 1.9e-246;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139
```

```
Qy       121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy       181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy       541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy       592 ------------------------------------------------------------ 591

Db       620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy       592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db       680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy       610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
            |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy       670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy      1030 TCGTTTACA 1038
            |||||||||
Db      1160 TCGTTTACA 1168
```

```
RESULT 7
HUMPCPBX
LOCUS       HUMPCPBX                1728 bp    mRNA    linear   PRI 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1  GI:189686
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1728)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
COMMENT     Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     gene            1. .1728
                     /gene="pCPB"
     CDS             20. .1291
                     /gene="pCPB"
                     /note="prepro-plasma carboxypeptidase B"
                     /codon_start=1
                     /protein_id="AAA60042.1"
                     /db_xref="GI:189687"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
     mat_peptide     86. .1288
                     /gene="pCPB"
                     /product="plasma carboxypeptidase B"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN
```

Query Match 84.2%; Score 909; DB 9; Length 1728;
Best Local Similarity 89.9%; Pred. No. 1.9e-246;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

```
Qy         1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy        61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy       121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy       181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        592 ------------------------------------------------------------ 591

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        592 -------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        |||||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1160 TCGTTTACA 1168


RESULT 8
I16100
LOCUS       I16100                  1749 bp    DNA     linear   PAT 03-APR-1996
DEFINITION  Sequence 2 from patent US 5474901.
ACCESSION   I16100
VERSION     I16100.1  GI:1251008
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Antibodies to human carboxypeptidase B and methods of use thereof
  JOURNAL   Patent: US 5474901-A 2 12-DEC-1995;

FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

  Query Match             84.2%;  Score 909;  DB 6;  Length 1749;
  Best Local Similarity   89.9%;  Pred. No. 1.9e-246;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 -------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        |||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000
```

```
Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 9
I33526
LOCUS       I33526                  1749 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 2 from patent US 5593674.
ACCESSION   I33526
VERSION     I33526.1  GI:1824317
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Plasma carboxypeptidase
  JOURNAL   Patent: US 5593674-A 2 14-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

  Query Match             84.2%;  Score 909;  DB 6;  Length 1749;
  Best Local Similarity   89.9%;  Pred. No. 1.9e-246;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 10
BT006936
LOCUS       BT006936                1272 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA,
            complete cds.
ACCESSION   BT006936
VERSION     BT006936.1  GI:30582710
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
```

```
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES              Location/Qualifiers
     source           1. .1272
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="GH00156X1.0"
                      /clone_lib="BD Creator(TM) CDS Library derived from MGC
                      collection"
                      /lab_host="DH5alpha T1 resistant"
                      /note="Vector: pDNR-Dual"
     CDS              1. .1272
                      /codon_start=1
                      /product="carboxypeptidase B2 (plasma, carboxypeptidase
                      U)"
                      /protein_id="AAP35582.1"
                      /db_xref="GI:30582711"
                      /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                      LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                      QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                      LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                      VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSCSE
                      TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                      LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                      GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      373 a    268 c    273 g    358 t
ORIGIN

  Query Match             84.0%;  Score 907.4;  DB 9;  Length 1272;
  Best Local Similarity   89.8%;  Pred. No. 5.3e-246;
  Matches 1032;  Conservative    0;  Mismatches    6;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        592 ------------------------------------------------------------ 591

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1141 TCGTTTACA 1149


RESULT 11
BT007990
LOCUS       BT007990               1272 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
            carboxypeptidase U) mRNA, partial cds.
ACCESSION   BT007990
VERSION     BT007990.1  GI:30584818
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
```

```
          Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00156L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .>1272
                     /note="Mutations: 1271:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens carboxypeptidase B2 (plasma,
                     carboxypeptidase U)"
                     /protein_id="AAP36662.1"
                     /db_xref="GI:30584819"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNVL"
BASE COUNT      372 a    268 c    273 g    359 t
ORIGIN

  Query Match             84.0%;  Score 907.4;  DB 12;  Length 1272;
  Best Local Similarity   89.8%;  Pred. No. 5.3e-246;
  Matches 1032;  Conservative    0;  Mismatches    6;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        592 ------------------------------------------------------------ 591

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1141 TCGTTTACA 1149


RESULT 12
BC007057
LOCUS       BC007057                1715 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
            IMAGE:3934520, mRNA, complete cds.
ACCESSION   BC007057
VERSION     BC007057.1  GI:13937896
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1715)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
```

```
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 16 Row: f Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4503004.
FEATURES              Location/Qualifiers
     source           1. .1715
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="LocusID:1361"
                      /db_xref="taxon:9606"
                      /clone="MGC:12495 IMAGE:3934520"
                      /tissue_type="Skeletal Muscle"
                      /clone_lib="NIH_MGC_81"
                      /lab_host="DH10B"
                      /note="Vector: pDNR-LIB"
     CDS              25. .1296
                      /codon_start=1
                      /product="carboxypeptidase B2 (plasma)"
                      /protein_id="AAH07057.1"
                      /db_xref="GI:13937897"
                      /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                      LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                      QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                      LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                      VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                      TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                      LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                      GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      518 a    348 c    341 g    508 t
ORIGIN

  Query Match             84.0%;  Score 907.4;  DB 9;  Length 1715;
  Best Local Similarity   89.8%;  Pred. No. 5.4e-246;
  Matches 1032;  Conservative    0;  Mismatches    6;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 84

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 144

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 204

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 264

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 324

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 384

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 444

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 504

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        505 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 564

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 624

Qy        592 ------------------------------------------------------------ 591

Db        625 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 684

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db        685 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 744

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        745 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 804

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        805 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 864

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 924

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 984

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1044

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1045 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1104

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1164

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1165 TCGTTTACA 1173


RESULT 13
AB021968
LOCUS       AB021968               1421 bp    mRNA    linear   ROD 27-JUL-2000
DEFINITION  Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION   AB021968
VERSION     AB021968.1  GI:9558447
KEYWORDS    carboxypeptidase R.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Obata,K., Okada,N.,
            Campbell,W. and Okada,H.
  TITLE     Pro-carboxypeptidase R is an acute phase protein in the mouse,
```

```
          whereas carboxypeptidase N is not
  JOURNAL   J. Immunol. 165 (2), 1053-1058 (2000)
  MEDLINE   20341711
   PUBMED   10878383
REFERENCE   2  (bases 1 to 1421)
  AUTHORS   Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1998) Tomoo Sato, Fukushimura Hospital, Choju
            Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
            441-8124, Japan (E-mail:tomoo@tcp-ip.or.jp, Tel:81-532-46-7511,
            Fax:81-532-46-4899)
FEATURES             Location/Qualifiers
     source          1. .1421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS             17. .1285
                     /codon_start=1
                     /product="carboxypeptidase R"
                     /protein_id="BAB03402.1"
                     /db_xref="GI:9558448"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
     polyA_site      1421
                     /note="11 a nucleotides"
BASE COUNT      418 a    320 c    305 g    378 t
ORIGIN

  Query Match           59.6%;  Score 643.4;  DB 10;  Length 1421;
  Best Local Similarity   76.1%;  Pred. No. 4e-171;
  Matches  874;  Conservative    0;  Mismatches  161;  Indels  114;  Gaps    2;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||    |||||| | |||| ||| |||    |  || ||| ||||||||||| |
Db         17 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCA---TCATCCTCTATGAGCAGCATGGC 73

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||| || |||||||||||| ||| || |||||||||| |||||||| ||||||||||||
Db         74 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 133

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              | || ||||||||||||||||| |||||| | |||||||||||||| || ||||||||
Db        134 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 193

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              | || | |||||||||  ||||||| |||||||| |||| |||||||||||||| |||
Db        194 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 253

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||| |||||||||||||||| |||||||||  |  ||  || ||    ||||||| ||
Db        254 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 313

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              || ||| ||||||||| || ||| ||||| |||||||||| |||||||| || ||||||
Db        314 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 373

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              || |||||||||| |||||||||||||||||| ||||||||| | ||||||||   ||||
Db        374 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 433

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||| |||||    |||||| |||| ||||| ||||| |||||||||||||| ||||||
Db        434 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 493

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||| || ||||| ||||||| |  |||||||||||| ||||| ||||||||||||
Db        494 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 553

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGC------------ 588
              |||||||||||||||| || |||||||||| ||||||||||||||
Db        554 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 613

Qy        589 ------------------------------------------------------------ 588

Db        614 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 673

Qy        589 ------------------------------------CATAATCGAATGTGGAGAAAG 609
                                                   | ||||||||||||| |||
Db        674 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 733

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||| |||    |  |||||| |  ||| | || |||||||||| || ||||| |
Db        734 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 793

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||| ||||||| || |||||||| ||||| ||||||||||||||
Db        794 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 853

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||| || ||||| |||||||||||||||    ||||||||||||||||
Db        854 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 913

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              |||| |||||||| |||||||| ||||| |||||||||| || ||| ||||||| |||
Db        914 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 973

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              |||||||   |||| |||||||| |||||  || ||||||||||||||| ||||| ||||||
Db        974 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1033

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||| ||||| |  | || |||||| |||||||| |||||  |||  ||||||||||
Db       1034 GTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1093

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |  ||||| |||||||||||||||  |||||||||||||||||||||||||||||||||||
Db       1094 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1153

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1154 TCGTTTACA 1162


RESULT 14
AF164524
LOCUS       AF164524               1490 bp    mRNA    linear   ROD 05-APR-2000
DEFINITION  Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
            mRNA, complete cds.
ACCESSION   AF164524
VERSION     AF164524.1  GI:7416966
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1490)
  AUTHORS   Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van
            Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
  TITLE     Characterization of mouse thrombin-activatable fibrinolysis
            inhibitor
  JOURNAL   Thromb. Haemost. 83 (2), 297-303 (2000)
  MEDLINE   20201996
   PUBMED   10739389
REFERENCE   2  (bases 1 to 1490)
  AUTHORS   Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
            Utrecht 3584 CX, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .1490
```

```
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:477035"
     gene            1. .1490
                     /gene="Tafi"
     CDS             84. .1352
                     /gene="Tafi"
                     /function="carboxypeptidase"
                     /codon_start=1
                     /product="thrombin-activatable fibrinolysis inhibitor"
                     /protein_id="AAF62385.1"
                     /db_xref="GI:7416967"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      438 a    337 c    322 g    393 t
ORIGIN
```

Query Match 59.6%; Score 643.4; DB 10; Length 1490;
Best Local Similarity 76.1%; Pred. No. 4e-171;
Matches 874; Conservative 0; Mismatches 161; Indels 114; Gaps 2;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||    ||||||  |  ||||  |||  |||    |   ||  |||  ||||||||||| |
Db         84 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCA---TCATCCTCTATGAGCAGCATGGC 140

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||  ||  ||||||||||||  |||  ||  ||||||||||  ||||||||  ||||||||||||
Db        141 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 200

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
                |  ||  ||||||||||||||||||  ||||||  |  ||||||||||||||  ||  ||||||||
Db        201 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 260

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
                |  ||  |  |||||||||   |||||||  ||||||||  |||||  ||||||||||||||  |||
Db        261 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 320

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||  ||||||||||||||||  |||||||||    |   ||   ||  ||     |||||||  ||
Db        321 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 380

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||  |||  |||||||||  ||  |||  |||||  ||||||||||  ||||||||  ||  ||||||
Db        381 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 440

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||  ||||||||||| ||||||||||||||||||  |||||||||  |  ||||||||    ||||
Db        441 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 500

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              |||||  |||||     ||||||  ||||  |||||  |||||  ||||||||||||||  ||||||
Db        501 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 560

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||  ||  |||||  |||||||  |  ||||||||||||  |||||  ||||||||||||
Db        561 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 620

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGC------------ 588
              ||||||||||||||||  ||  ||||||||||  ||||||||||||||
Db        621 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 680

Qy        589 ---------------------------------------------------------- 588

Db        681 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 740

Qy        589 ----------------------------------------CATAATCGAATGTGGAGAAAG 609
                                                       |  ||||||||||||| |||
Db        741 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 800

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||  |||    |   ||||||  |  |||  |  ||  ||||||||||  ||  |||||  |
Db        801 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 860

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||  |||||||  ||  ||||||||  |||||  ||||||||||||||||
Db        861 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 920

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||  ||  |||||  ||||||||||||||||    ||||||||||||||||||
Db        921 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 980

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||  ||||||||  ||||||||  |||||  ||||||||||  ||  |||  |||||||  |||
Db        981 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 1040

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              |||||||   ||||  ||||||||  |||||  ||  ||||||||||||||  |||||  ||||||
Db       1041 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1100

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||  |||||  |  |  ||  ||||||  ||||||||  |||||  |||  ||||||||||
Db       1101 GTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1160

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |  |||||  ||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db       1161 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1220

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1221 TCGTTTACA 1229
```

```
RESULT 15
AB042598
LOCUS       AB042598                1425 bp    mRNA    linear   ROD 02-DEC-2000
DEFINITION  Rattus norvegicus mRNA for pre-procarboxypeptidase R, complete cds.
ACCESSION   AB042598
VERSION     AB042598.1  GI:11526576
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (sites)
  AUTHORS   Kato,T., Akatsu,H., Sato,T., Matsuo,S., Yamamoto,T., Campbell,W.,
            Hotta,N., Okada,N. and Okada,H.
  TITLE     Molecular cloning and partial characterization of rat
            procarboxypeptidase R and carboxypeptidase N
  JOURNAL   Microbiol. Immunol. 44 (8), 719-728 (2000)
  MEDLINE   20471387
   PUBMED   11021404
REFERENCE   2  (bases 1 to 1425)
  AUTHORS   Kato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2000) Tomomi Kato, Choju Medical Institute, Choju
            Medical Institute; Noyori-cho, Yamanaka, 19-14, Toyohashi, Aichi
            441-8124, Japan (E-mail:katotomo@tsuru.med.nagoya-u.ac.jp,
            Tel:0532-46-7511, Fax:0532-46-4899)
FEATURES             Location/Qualifiers
     source          1. .1425
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10116"
     CDS             100. .1368
                     /codon_start=1
                     /product="pre-procarboxypeptidase R"
                     /protein_id="BAB18617.1"
                     /db_xref="GI:11526577"
```

```
                     /translation="MKLYGLGVLVAIILYEKHGLAFQSGHVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVNSVKAYLNASRIPFNVLMNNVEDLIQ
                     QQTSNDTVSPRASSSYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSYEKYPLYVL
                     KVSGKEHRVKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENTYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSVHMNNRCVGTDLNRNFASKHWCEKGASSFSCSET
                     YCGLYPESEPEVKAVADFLRRNINHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERFIKPTCAEALAAVSKIAWHVIRNS"
BASE COUNT      417 a    334 c    309 g    365 t
ORIGIN

  Query Match           59.5%;  Score 642.4;  DB 10;  Length 1425;
  Best Local Similarity  76.0%;  Pred. No. 7.6e-171;
  Matches  873;  Conservative    0;  Mismatches  161;  Indels  114;  Gaps    2;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||   ||||| |||||| ||| |||    |  || ||| |||| |||||| |
Db        100 ATGAAGCTTTATGGCCTTGGAGTCCTGGTAGCCA---TCATCCTCTATGAGAAGCATGGC 156

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              | || || |||||||||||| |||||| ||||||| ||| ||||||| ||||||||||||
Db        157 CTTGCCTTTCAGAGTGGCCATGTTCTATCTGCTCTCCCTCGAACCTCCAGGCAAGTTCAA 216

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              |||| |||||||| |||||||| || ||| ||||||||||||||||| || ||||||||
Db        217 CTTCTTCAGAATCTCACTACAACTTACGAGGTTGTTCTCTGGCAGCCAGTGACAGCTGAA 276

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              | |||| |||||||||  ||||||| || ||||| ||||| |||||||||| ||| |||
Db        277 TTCATTGAGAAGAAAAAGGAAGTCCACTTCTTTGTGAATGCGTCTGATGTCAACAGTGTC 336

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||| ||||||||| |||| |||||||||  |  ||| || ||    ||||||| |||
Db        337 AAAGCCTATTTAAATGCGAGCAGAATTCCATTTAACGTCCTGATGAACAACGTGGAGGAT 396

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              || |||||||||||||  ||||| ||||| || ||||||||||||||| | || ||||||
Db        397 CTAATTCAACAGCAGACGTCCAATGACACTGTTAGCCCCCGAGCCTCCTCCTCATACTAT 456

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              |||||||||||||  |||||||||||||||| ||||||||| ||||||||||   |||
Db        457 GAACAGTATCACTCGTTAAATGAAATCTATTCCTGGATAGAAGTTATAACTGAACAGCAC 516

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||| |||||    |||||| |||||||||||||| ||| ||||||||||| |||||
Db        517 CCTGACATGCTCCAGAAAATCTACATTGGATCCTCATATGAAAAGTACCCACTTTATGTG 576

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||| || ||||| ||||| | || |||||||||||||||||| ||||||||||||
Db        577 TTAAAGGTCTCAGGAAAGGAACACAGAGTCAAAAATGCCATATGGATCGACTGTGGAATC 636

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||  ||||| || ||  ||||||||||||||||||||||||  ||
Db        637 CATGCCAGAGAGTGGATTTCACCAGCTTTCTGCTTGTGGTTCATAGGCTATGTAACGCAA 696

Qy        592 ------------------------------------------------------------ 591

Db        697 TTCCATGGGAAAGAAAATACATACACCAGACTTCTGAGGCACGTGGATTTCTACATTATG 756

Qy        592 -------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        757 CCAGTGATGAATGTGGACGGCTACGACTACACGTGGAAAAAGAATCGAATGTGGAGAAAG 816

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||| ||| || |   |||||| |  ||| | || |||||||||| || ||||| |
Db        817 AACCGCTCTGTCCACATGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 876

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||| |||| ||||| |||| ||| ||||| || ||||||||||||
Db        877 TCCAAACACTGGTGTGAGAAAGGCGCATCAAGTTTCTCCTGCTCTGAGACCTACTGTGGA 936

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||| |||||||| || ||||| |||||||||||||||   ||| |||| |||||||||

Db        937 CTTTACCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCCTGAGGAGAAATATC 996

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||| |||||||| |||||||| ||||| |||||||||||||| ||| ||||||| |||
Db        997 AACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAGCAAATACTGTTTCCCTAT 1056

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||| |   |||| |||||||| ||||| |||||||||||||||| ||||| ||||||
Db       1057 TCCTACAACAGAAGCAAAAGCAAGGACCACGAGGAACTGTCTCTAGTGGCCAGCGAAGCA 1116

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||| ||||| |  | || |||||| |||||||| |||||||||  |||||||||
Db       1117 GTTCGTGCCATTGAAAGTATTAATAAAAACACCAGGTACACACATGGCAGTGGCTCAGAA 1176

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |  ||||| ||||||||||||||||   || ||||||||||||||||||||||||||||
Db       1177 AGTTTATATCTAGCTCCTGGAGGTTCTGATGATTGGATCTATGATTTGGGCATCAAATAT 1236

Qy       1030 TCGTTTAC 1037
              ||||||||
Db       1237 TCGTTTAC 1244


Search completed: January  1, 2004, 13:55:55
Job time : 4083.16 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 09:40:19 ; Search time 317.12 Seconds
(without alignments)
9193.336 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1080	100.0	1573	21	AAC81962	Human brain carbox
2	1025.2	94.9	1625	25	ABX13670	Human protease cDN
3	909	84.2	1272	20	AAV74302	Human plasma carbo
4	909	84.2	1728	24	ABN95899	Gene #2397 used to
5	909	84.2	1749	17	AAT11671	Human plasma carbo
6	909	84.2	1749	18	AAT62846	Human plasma carbo
7	907.4	84.0	1749	14	AAQ41001	Human plasma carbo
8	441	40.8	1400	21	AAF18005	Lung cancer associ
9	331.2	30.7	416	25	ABX49430	Bovine EST associa
10	282.4	26.1	397	25	ABX43260	Bovine EST associa
11	204	18.9	55827	25	ABX13671	Human protease gen
12	167	15.5	231	24	ABN19133	Human ORFX polynuc
c 13	129.6	12.0	431	24	ABK62561	Rat sequence diffe
14	124	11.5	927	17	AAT35760	Rat mature carboxy
15	123.4	11.4	921	16	AAQ90601	Porcine carboxypep
16	123.4	11.4	1215	16	AAQ90600	Porcine Tyr-His-Me
17	109	10.1	1263	19	AAV41795	Human pancreatic c
18	107.6	10.0	2023	25	ABZ75116	Anti-human seminal
19	107.4	9.9	999	17	AAT42500	Mature HCPB coding
20	107.4	9.9	1053	17	AAT42497	mature HCPB-(His)6
21	107.4	9.9	1263	17	AAT42494	Human pancreatic c
22	107.4	9.9	1284	17	AAT42506	ProHCPB gene with
23	104.2	9.6	1053	20	AAX24804	Human carboxypepta
24	104.2	9.6	1059	17	AAT42511	Modified HCPB (D25
25	104.2	9.6	1059	17	AAT42512	Modified HCPB (D25
26	104.2	9.6	1059	18	AAT62787	Carboxypeptidase B
27	104.2	9.6	1059	18	AAT62788	Carboxypeptidase B
28	102.6	9.5	1059	18	AAT62805	Carboxypeptidase B
29	102.6	9.5	1059	18	AAT62806	Carboxypeptidase B
30	102.6	9.5	1059	18	AAT62791	Carboxypeptidase B
31	102.6	9.5	1059	18	AAT62801	Carboxypeptidase B
32	101	9.4	1059	18	AAT62789	Carboxypeptidase B
33	101	9.4	1059	18	AAT62790	Carboxypeptidase B
34	101	9.4	1870	18	AAV17322	DNA encoding human
35	101	9.4	1870	19	AAV41819	Human carboxypepti
36	101	9.4	2154	18	AAV17331	PreproHCPB-linker-
37	99.4	9.2	1059	18	AAT62808	Carboxypeptidase B
38	99.4	9.2	1059	18	AAT62804	Carboxypeptidase B
39	99.2	9.2	1059	18	AAT62800	Carboxypeptidase B
40	97.8	9.1	1059	18	AAT62807	Carboxypeptidase B
41	97.8	9.1	1059	18	AAT62803	Carboxypeptidase B
42	96.2	8.9	1059	18	AAT62802	Carboxypeptidase B
43	94.6	8.8	1332	24	ABL65831	Lung cancer relate
44	92.8	8.6	1302	25	ABV77254	Nucleotide sequenc
45	92.8	8.6	1311	22	AAD03837	Human carboxypepti

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.
XX
AC AAC81962;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B cDNA.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 18..1100
FT /*tag= a
FT /product= "carboxypeptidase B"
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX

```
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
DR   P-PSDB; AAB11457.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Claim 2b; Page 64-68; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

  Query Match             100.0%;  Score 1080;  DB 21;  Length 1573;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1080;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037

Qy       1021 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097

RESULT 2
ABX13670
ID   ABX13670 standard; cDNA; 1625 BP.
XX
AC   ABX13670;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease cDNA.
XX
KW   Human; gene; ss; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..16
FT                   /*tag=  a
FT   CDS             17..1177
FT                   /*tag=  b
FT                   /product= "Protease"
FT   variation       replace(521,A)
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR           1178..1625
FT                   /*tag=  d
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
```

```
XX
DR   WPI; 2003-102387/09.
DR   P-PSDB; ABG72093, ABG72095.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 4; Fig 1; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   cDNA, the gene for which is located on chromosome 13.
XX
SQ   Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

  Query Match           94.9%;  Score 1025.2;  DB 25;  Length 1625;
  Best Local Similarity  99.2%;  Pred. No. 8.5e-311;
  Matches 1030;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 316

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 556

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 616

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 676

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              |||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 736

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 796

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 856

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db        917 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 976

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1036

Qy       1021 ATCAAATATTCGTTTACA 1038
              ||||||||||||||||||
Db       1037 ATCAAATATTCGTTTACA 1054


RESULT 3
AAV74302
ID   AAV74302 standard; cDNA; 1272 BP.
XX
AC   AAV74302;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW   polymorphism detection; thrombotic disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   P-PSDB; AAW92270.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
```

```
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Example 1; Page 24; 35pp; English.
XX
CC   This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

  Query Match             84.2%;  Score 909;  DB 20;  Length 1272;
  Best Local Similarity   89.9%;  Pred. No. 2.2e-274;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||.||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        592 ------------------------------------------------------------ 591

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        592 ---------------------------------------------AATCGAATGTGGAGAAAG 609
                                                          ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1141 TCGTTTACA 1149


RESULT 4
ABN95899
ID   ABN95899 standard; DNA; 1728 BP.
XX
AC   ABN95899;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2397 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 2397; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
```

```
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;

  Query Match             84.2%;  Score 909;  DB 24;  Length 1728;
  Best Local Similarity   89.9%;  Pred. No. 2.6e-274;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        592 ------------------------------------------------------------ 591

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1160 TCGTTTACA 1168


RESULT 5
AAT11671
ID   AAT11671 standard; DNA; 1749 BP.
XX
AC   AAT11671;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human plasma carboxypeptidase B.
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   P-PSDB; AAR90293.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
```

```
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

  Query Match             84.2%;  Score 909;  DB 17;  Length 1749;
  Best Local Similarity   89.9%;  Pred. No. 2.6e-274;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 ---------------------------------------------AATCGAATGTGGAGAAAG 609
                                                          ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 6
AAT62846
ID   AAT62846 standard; DNA; 1749 BP.
XX
AC   AAT62846;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human PCPB
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
FT   misc_binding    134..177
FT                   /*tag=  d
FT                   /bound_moeity= 46_bp_probe
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
```

```
DR   P-PSDB; AAW14733.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence encodes human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

  Query Match            84.2%;  Score 909;  DB 18;  Length 1749;
  Best Local Similarity  89.9%;  Pred. No. 2.6e-274;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 -----------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 7
AAQ41001
ID   AAQ41001 standard; cDNA; 1749 BP.
XX
AC   AAQ41001;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B gene.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     41..106
FT                   /*tag=  a
FT   mat_peptide     107..1312
FT                   /*tag=  b
FT                   /product= PCPB
FT   misc_feature    134..177
FT                   /*tag=  c
FT                   /function= probe
FT                   /note= "used to obtain full-length clones"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   P-PSDB; AAR36273.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
```

```
PS   Disclosure; Fig 4; 40pp; English.
XX
CC   Human plasma carboxypeptidase B was isolated from human plasma and
CC   partially sequenced. Oligonucleotide primers were designed based on
CC   the partial amino acid sequences. The primers were used in a PCR
CC   amplification to identify cDNA encoding PCPB from a human liver cDNA
CC   library. The PCR product was capable of encoding the first 37 amino
CC   acids of PCPB; a 46mer probe was used to obtain the full-length
CC   sequence which, although disclosed in the specification, is not
CC   claimed. PCPB inhibits the enzymatic conversion by tPA of
CC   plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

  Query Match             84.0%;  Score 907.4;  DB 14;  Length 1749;
  Best Local Similarity   89.8%;  Pred. No. 8.3e-274;
  Matches 1032;  Conservative    0;  Mismatches    6;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 -----------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 8
AAF18005
ID   AAF18005 standard; DNA; 1400 BP.
XX
AC   AAF18005;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58129.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 1; Page 507; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
```

```
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

  Query Match            40.8%;  Score 441;  DB 21;  Length 1400;
  Best Local Similarity  98.9%;  Pred. No. 1.8e-127;
  Matches  444;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy        590 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 649
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 AGAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 236

Qy        650 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 709
              |||||||||| |||||||  ||||||||||||||||||||||||||||||||||||||||
Db        237 ACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 296

Qy        710 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 769
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 356

Qy        770 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 416

Qy        830 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 476

Qy        890 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATA 949
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        477 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATA 536

Qy        950 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 596

Qy       1010 ATGATTTGGGCATCAAATATTCGTTTACA 1038
              |||||||||||||||||||||||||||||
Db        597 ATGATTTGGGCATCAAATATTCGTTTACA 625


RESULT 9
ABX49430
ID   ABX49430 standard; cDNA; 416 BP.
XX
AC   ABX49430;
XX
DT   21-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #14595.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;   99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle  -
XX
PS   Claim 2; SEQ ID No 14595; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ   Sequence 416 BP; 125 A; 86 C; 93 G; 112 T; 0 other;

  Query Match            30.7%;  Score 331.2;  DB 25;  Length 416;
  Best Local Similarity  87.3%;  Pred. No. 2.7e-93;
  Matches  363;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy        150 GATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTT 209
              |||||||||||||||||| ||| |||||||   |||||||||  | |  ||||||||||
Db          1 GATTGTTCTCTGGCAGCCAGTAGCAGCTGAATATATTGTGAAGGGATACGAAGTCCATTT 60

Qy        210 TTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCC 269
              |||||| ||||||||||||||   ||||||||||||||||||||||| |||| |||||||
Db         61 TTTTGTGAATGCATCTGATGTAAGCAATGTGAAAGCCCATTTAAATGCGAGCAGAATTCC 120

Qy        270 ATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACAC 329
              || ||| ||| || ||| | | |||||||||||||| |  ||||||| |||||| |||||
Db        121 ATTCAGGGTCCTGGTGGAAAATGTGGAAGATCTTATCCGGCAGCAGACTTCCAATGACAC 180

Qy        330 AGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTA 389
              |||||||||| || ||| | || |||||||||||||||||||||||||||||| |||||
Db        181 CATCAGCCCCCGGGCATCCTCCTCCTACTATGAACAGTATCACTCACTAAATGAGATCTA 240

Qy        390 TTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGG 449
              |||||||||||| |||| |||||| || |||||||||| ||  ||||||||||||||||
Db        241 TTCTTGGATAGAAGTTATGACTGAGCGGTATCCTGATATGGTTGAAAAAATCCACATTGG 300

Qy        450 ATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGC 509
```

```
Db        301 ATCCTCATACGAGAAGTACCCACTTTATGTTTTAAAGGTTTCTAAAAAAGAACAAAGGGC 360

Qy        510 CAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTG 565

Db        361 CAAAAATGCCATGTGGATTGACTGTGGAATCCACGCCAGAGAGTGGATCTCTCCTG 416


RESULT 10
ABX43260
ID   ABX43260 standard; cDNA; 397 BP.
XX
AC   ABX43260;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #8425.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-0960352.
XX
PR   12-JAN-1999;   99US-115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC,  Mathialagan N,  Tao N,  Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and
PT   analysis, cattle breeding, or for genetically improving cattle  -
XX
PS   Claim 2; SEQ ID No 8425; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived
CC   from cattle, and the LMFD nucleic acid can specifically hybridise to a
CC   second nucleic acid molecule comprising any of 15112 nucleotide
CC   sequences, appearing as ABX34836-ABX49947, or complements of them.
CC   Also included are; (1) a transformed cell having a nucleic acid
CC   comprising an LMFD nucleic acid linked to a promoter and a 3' non-
CC   translated sequence that functions in the cell to cause termination of
CC   transcription and addition of polyadenylated ribonucleotides to a 3' end
CC   of the mRNA molecule; and (2) determining a level or pattern of a
CC   molecule in a bovine cell or tissue comprising: (a) incubating a marker
CC   nucleic acid (comprising any of the 15112 nucleic acid sequences or its
CC   complement or fragment) with a complementary nucleic acid molecule
CC   obtained from the bovine cell or tissue, where hybridisation between the
CC   marker nucleic acid and the complementary nucleic acid permits the
CC   detection of the molecule; and (b) detecting the level or pattern of the
CC   complementary nucleic acid, where the detection of the complementary
CC   nucleic acid is predictive of the level or pattern of the molecule.
CC   The LMFD nucleic acid is used for determining a level or pattern
CC   of a molecule in a bovine cell or tissue. It is useful for genome
CC   mapping, gene identification and analysis, cattle breeding, preparation
CC   of constructs for use in cattle gene expression, or for genetically
CC   improving cattle. The present sequence is one of the 15112 bovine
CC   LMFD EST (expressed sequence tag) nucleic acids.
CC   Note: The present sequence was not shown in the specification but
CC   was obtained in electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139.
XX
SQ   Sequence 397 BP; 100 A; 94 C; 96 G; 107 T; 0 other;

  Query Match             26.1%;  Score 282.4;  DB 25;  Length 397;
  Best Local Similarity   83.9%;  Pred. No. 5.4e-78;
  Matches  319;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Db         18 ATGAAGCTTTATAGCTTGGGAGTCCTTGTCGCCACCGTCCTGTTCTGTGGGGAGCATGCC 77

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Db         78 TTCGCCTTTCAGAGGGGCCAGGTTTTATCTGCTCTTCCTAGAACCTCCAGGCAAGTTCAA 137

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Db        138 ATTCTGCAGAATGTTACTACAACATACAAGATTGTTCTCTGGCAGCCAGTAGCAGCTGAA 197

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Db        198 TATATTGTGAAGGGATACGAAGTCCATTTTTTTGTGAATGCATCTGATGTAAGCAATGTG 257

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Db        258 AAAGCCCATTTAAATGCGAGCAGAATTCCATTCAGGGTCCTGGTGGAAAATGTGGAAGAT 317

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Db        318 CTTATCCGGCAGCAGACTTCCAATGACACCATCAGCCCCCGGGCATCCTCCTCCTACTAT 377

Qy        361 GAACAGTATCACTCACTAAA 380

Db        378 GAACAGTATCACTCACTAAA 397


RESULT 11
ABX13671
ID   ABX13671 standard; DNA; 55827 BP.
XX
AC   ABX13671;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease gene.
XX
KW   Human; gene; ds; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             2017..53409
FT                   /*tag=  a
FT                   /product= "Protease"
FT   exon            2017..2090
FT                   /*tag=  b
FT                   /number= 1
FT   intron          2091..19206
FT                   /*tag=  c
FT                   /number= 1
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            19207..19282
FT                   /*tag=  d
FT                   /number= 2
FT   intron          19283..22683
FT                   /*tag=  e
FT                   /number= 2
```

```
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            22684..22808
FT                   /*tag= f
FT                   /number= 3
FT   intron          22809..24477
FT                   /*tag= g
FT                   /number= 3
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            24478..24586
FT                   /*tag= h
FT                   /number= 4
FT   intron          24587..28129
FT                   /*tag= i
FT                   /number= 4
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            28130..28231
FT                   /*tag= j
FT                   /number= 5
FT   intron          28232..33052
FT                   /*tag= k
FT                   /number= 5
FT   exon            33053..33157
FT                   /*tag= l
FT                   /number= 6
FT   intron          33158..42288
FT                   /*tag= m
FT                   /number= 6
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            42289..42382
FT                   /*tag= n
FT                   /number= 7
FT   intron          42383..48643
FT                   /*tag= o
FT                   /number= 7
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            48644..48846
FT                   /*tag= p
FT                   /number= 8
FT   intron          48847..51176
FT                   /*tag= q
FT                   /number= 8
FT                   /cons_splice= (5'site:yes,3'site:no)
FT   exon            51177..51264
FT                   /*tag= r
FT                   /number= 9
FT   intron          51265..53224
FT                   /*tag= s
FT                   /number= 9
FT                   /cons_splice= (5'site:no,3'site:no)
FT   exon            53225..53409
FT                   /*tag= t
FT                   /number= 10
FT   variation       replace(858,T)
FT                   /*tag= u
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(2122,T)
FT                   /*tag= v
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4088,A)
FT                   /*tag= w
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4240,A)
FT                   /*tag= x
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4416..4418,TA)
FT                   /*tag= y
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4434,G)
FT                   /*tag= z
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(4969,C)
FT                   /*tag= aa
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(5868,C)
FT                   /*tag= ab
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(6254,C)
FT                   /*tag= ac
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10171,C)
FT                   /*tag= ad
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(10452,A)
FT                   /*tag= ae
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(11613,A)
FT                   /*tag= af
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(12130,T)
FT                   /*tag= ag
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(17867..17868,GAA)
FT                   /*tag= ah
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(18243,G)
FT                   /*tag= ai
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(22450,C)
FT                   /*tag= aj
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(23003,T)
FT                   /*tag= ak
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24055..24056,GAA)
FT                   /*tag= al
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(24132,C)
FT                   /*tag= am
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(25713,A)
FT                   /*tag= an
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(28476,C)
FT                   /*tag= ao
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(29404,G)
FT                   /*tag= ap
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31047,T)
FT                   /*tag= aq
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31445,C)
FT                   /*tag= ar
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31447,C)
FT                   /*tag= as
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31600,C)
FT                   /*tag= at
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31714,C)
FT                   /*tag= au
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(31715,T)
FT                   /*tag= av
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32193,G)
FT                   /*tag= aw
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32341,G,A)
FT                   /*tag= ax
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32561,G)
FT                   /*tag= ay
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32600..35601,AAA)
```

```
FT                   /*tag=  az
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32642,T)
FT                   /*tag=  ba
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(32793..32794,TAG)
FT                   /*tag=  bb
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(33071,A)
FT                   /*tag=  bc
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(34721,T)
FT                   /*tag=  bd
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35304,A)
FT                   /*tag=  be
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(35425,C)
FT                   /*tag=  bf
FT                   /standard_name= "Single nucleotide polymorphism"
FT   variation       replace(36050,G)

  Query Match             18.9%;  Score 204;  DB 25;  Length 55827;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-52;
  Matches  204;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42043 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 42102

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42103 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 42162

Qy        805 TACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGT 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42163 TACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGT 42222

Qy        865 AAAAGCAAAGACCATGAGGAACTG 888
              ||||||||||||||||||||||||
Db      42223 AAAAGCAAAGACCATGAGGAACTG 42246


RESULT 12
ABN19133
ID   ABN19133 standard; cDNA; 231 BP.
XX
AC   ABN19133;
XX
DT   24-JUN-2002  (first entry)
XX
DE   Human ORFX polynucleotide sequence SEQ ID NO:6743.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   P-PSDB; ABP03381.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders  -
XX
PS   Disclosure; SEQ ID 6743; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, neurodegenerative disorders, disorders related to organ
CC   transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC   useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC   bone degenerative disorders, or periodontal disease, and for gut
CC   protection or regeneration and treatment of lung or liver fibrosis,
CC   reperfusion injury in various tissues and conditions resulting from
CC   systemic cytokine damage.
CC   N.B.  The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 231 BP; 74 A; 58 C; 40 G; 59 T; 0 other;

  Query Match             15.5%;  Score 167;  DB 24;  Length 231;
  Best Local Similarity   82.7%;  Pred. No. 6.7e-42;
  Matches  191;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy        223 TCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTG 282
              |||||||||  |||  |||  ||||||  |||||||||  ||||  |||||||||   |  ||| ||
Db          1 TCTGATGTCAACAGTGTCAAAGCCTATTTAAATGCGAGCAGAATTCCATTTAACGTCCTG 60

Qy        283 CTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGA 342
              ||    |||||||  |||||  ||||||||||||||  |||||  |||||  ||  |||||||||
Db         61 ATGAACAACGTGGAGGATCTAATTCAACAGCAGACGTCCAATGACACTGTTAGCCCCCGA 120

Qy        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402
              ||||||  |  ||  ||||||||||||||||||||  ||||||||||||||||  |||||||||
Db        121 GCCTCCTCCTCATACTATGAACAGTATCACTCGTTAAATGAAATCTATTCCTGGATAGAA 180

Qy        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 453
              ||||||||||    |||  |||||  |||||     ||||||  ||||||||||
Db        181 GTTATAACTGAACAGCACCCTGACATGCTCCAGAAAATCTACATTGGATCC 231


RESULT 13
ABK62561/c
ID   ABK62561 standard; cDNA; 431 BP.
XX
AC   ABK62561;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Rat sequence differentially expressed in response to a hepatotoxin #468.
XX
KW   Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
```

```
KW   differential expression; centrilobular necrosis; steatosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200210453-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-US23872.
XX
PR   31-JUL-2000; 2000US-222040P.
PR   02-NOV-2000; 2000US-244880P.
PR   11-MAY-2001; 2001US-290029P.
PR   15-MAY-2001; 2001US-290645P.
PR   22-MAY-2001; 2001US-292336P.
PR   06-JUN-2001; 2001US-295798P.
PR   13-JUN-2001; 2001US-297457P.
PR   19-JUN-2001; 2001US-298884P.
PR   09-JUL-2001; 2001US-303459P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D,  Porter MW,  Johnson KR,  Castle AL,  Elashoff MR;
XX
DR   WPI; 2002-241625/29.
XX
PT   Predicting toxic effects of compounds or the progression of these toxic
PT   effects by determining the changes in gene expression in tissues or
PT   cells exposed to the toxin and comparing these to gene expression in
PT   unexposed tissues or cells -
XX
PS   Claim 1; Seq ID No 468; 239pp; English.
XX
CC   The invention relates to methods for predicting toxic effects of
CC   compounds or the progression of these toxic effects by determining the
CC   global changes in gene expression in tissues or cells exposed to the
CC   toxin and comparing these to gene expression in unexposed tissues or
CC   cells. Also included are methods of predicting at least one toxic
CC   effect of a compound or progression of a toxic effect, preferably the
CC   hepatotoxicity of a compound, comprising detecting the level of
CC   expression in a tissue or cell sample exposed to the compound of two or
CC   more genes listed in the specification, where  differential expression of
CC   the genes is indicative of at least one toxic effect or progression.
CC   The method can also be used to identify an agent which modulates the
CC   toxic response and predict cellular pathways that a compound modulates
CC   in a cell. The methods utilise a set of at least two probes (on a solid
CC   support in kit form), where each of the probes comprises a sequence that
CC   specifically hybridises to a gene listed in the specification, a computer
CC   system comprising a database containing information identifying the
CC   expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC   set of genes comprising at least two genes listed in the specification,
CC   and a user interface to view the information used to present information
CC   identifying the expression level in a tissue or cell of at least one gene
CC   listed in the specification. The method is useful for elucidating global
CC   changes in gene expression and for identifying toxicity markers in
CC   tissues or cell exposed to a known toxin. The genes may be used as
CC   toxicity markers in drug screening and toxicity assays. The genes and
CC   gene expression information may be used as diagnostic markers for the
CC   prediction or identification of the physiological state of tissue or cell
CC   sample that has been exposed to a compound or agent. Hepatotoxicity
CC   is characterised by centrilobular necrosis and steatosis. The present
CC   sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC   which is differentially expressed in response to a hepatotoxic agent.
XX
SQ   Sequence 431 BP; 130 A; 85 C; 84 G; 131 T; 1 other;

  Query Match           12.0%;  Score 129.6;  DB 24;  Length 431;
  Best Local Similarity  87.6%;  Pred. No. 5.1e-30;
  Matches  141;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        877 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 936
              ||||||||||||||| |||| ||||| |||||||||||||| ||||| |  | || ||||
Db        431 CATGAGGAACTGTCTNTAGTGGCCAGCGAAGCAGTTCGTGCCATTGAAAGTATTAATAAA 372

Qy        937 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 996
              || |||||||| |||||||||  |||||||||||  ||||| ||||||||||||||||
Db        371 AACACCAGGTACACACATGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCT 312

Qy        997 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC 1037
              || |||||||||||||||||||||||||||||||||||||
Db        311 GATGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC 271


RESULT 14
AAT35760
ID   AAT35760 standard; cDNA; 927 BP.
XX
AC   AAT35760;
XX
DT   09-NOV-1996  (first entry)
XX
DE   Rat mature carboxypeptidase B cDNA.
XX
KW   Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (1..50)
FT                   /*tag=  a
FT                   /note= "Mature CPB 5'-end primer"
FT   primer_bind     898..927
FT                   /*tag=  b
FT                   /note= "CPB 3' end primer"
XX
PN   WO9623064-A1.
XX
PD   01-AUG-1996.
XX
PF   25-JAN-1996;   96WO-US00995.
XX
PR   25-JAN-1995;   95US-0378233.
XX
PA   (BIOT-) BIOTECHNOLOGY GEN CORP.
XX
PI   Fulga N,  Gorecki M,  Hartman J,  Mendelovitch S;
XX
DR   WPI; 1996-362688/36.
DR   P-PSDB; AAW00602.
XX
PT   Purified active recombinant carboxypeptidase B prodn. - by
PT   expressing DNA encoding the pro-enzyme, folding and enzymatic
PT   cleavage to give active form, used e.g. for insulin prodn.
XX
PS   Example 1; Page 37-38; 49pp; English.
XX
CC   A cDNA sequence (AAT35760) codes for the mature form (AAW00602) of
CC   rat carboxypeptidase B (CPB).  A full-length sequence including
CC   cDNA (AAT35759) coding for the activation peptide (AAW00601) can
CC   be obtd. by PCR amplification (see also AAT35756 and AAT35758) of cDNA
CC   derived from Sprague-Dawley rat pancreas.  Pro-CPB can be expressed
CC   in host cells, e.g. Escherichia coli, and subsequently recovered,
CC   refolded and cleaved with trypsin to yield the active enzyme.  CPB
CC   produced this way is cheaper than porcine pancreatic enzyme, and is
CC   free of other proteases.
XX
SQ   Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

  Query Match           11.5%;  Score 124;  DB 17;  Length 927;
  Best Local Similarity  57.9%;  Pred. No. 4.4e-28;
  Matches  278;  Conservative    0;  Mismatches  195;  Indels    7;  Gaps    3;

Qy        596 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 655
              |||||||||||| | ||| |||       ||    | |   ||| | ||   |||||  |
Db        362 GAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGACCCCA 421
```

```
Qy        656 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 715
Db        422 ACAGGAATTTTAATGC---TGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478

Qy        716 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 775
Db        479 AAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGATTTCA 538

Qy        776 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 835
Db        539 TCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAGATGA 598

Qy        836 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 895
Db        599 TGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAATGCCC 658

Qy        896 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 955
Db        659 TGGTGAAAGGTGCGGCAAAGGAGCTTGCC---ACTCTGCATGGCACCAAGTACACATATG 715

Qy        956 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1015
Db        716 GCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTATGATC 775

Qy       1016 TGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTC 1075
Db        776 AGGGAATCAAATATTCCTTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834
```

RESULT 15
AAQ90601
ID AAQ90601 standard; DNA; 921 BP.
XX
AC AAQ90601;
XX
DT 13-MAR-1996 (first entry)
XX
DE Porcine carboxypeptidase B coding sequence.
XX
KW Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW human serum albumin; premating factor alpha; mating factor alpha;
KW proCBB; ds.
XX
OS Sus scrofa.
XX
FH Key Location/Qualifiers
FT CDS 1..921
FT /*tag= a
FT /product= Porcine carboxypeptidase B.
XX
PN WO9514096-A1.
XX
PD 26-MAY-1995.
XX
PF 16-NOV-1994; 94WO-US13142.
XX
PR 16-NOV-1993; 93US-0153258.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Fayerman JT, Greenen DP, Hershberger CL, Larson JL;
PI Sterner JL, Zhang H;
XX
DR WPI; 1995-200386/26.
DR P-PSDB; AAR75132.
XX
PT DNA encoding porcine carboxy:peptidase B - used for transforming
PT host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS Claim 1; Page 20; 34pp; English.
XX
CC The porcine carboxypeptidase B coding sequence can be place in a
CC bacterial or pref. Pichia yeast expression vector. The expression
CC vector further comprises the signal peptide of either human serum
CC albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC (designated pLGD27 - NRRL B-21027). The method can be used for
CC producing large amounts of porcine carboxypeptidase B and when
CC produced in Pichia yeast, the protein does not need solubilisation
CC or folding. The produced enzyme is then used for pref. cleaving
CC basic residues from the carboxy terminus of proteins.
XX
SQ Sequence 921 BP; 270 A; 238 C; 194 G; 219 T; 0 other;

Query Match 11.4%; Score 123.4; DB 16; Length 921;
Best Local Similarity 56.3%; Pred. No. 6.8e-28;
Matches 273; Conservative 0; Mismatches 206; Indels 6; Gaps 2;

```
Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
Db        321 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCG 380

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
Db        381 CTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTTGATGCTGG 440

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
Db        441 G---TGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGC 497

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
Db        498 TGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTC 557

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
Db        558 CATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTA 617

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
Db        618 TGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAA 677

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
Db        678 AGAACTTG---CTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAAT 734

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
Db        735 CTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTT 794

Qy       1035 TACAT 1039
Db        795 CACCT 799
```

Search completed: January 1, 2004, 11:08:45
Job time : 325.12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:34:47 ; Search time 5559.33 Seconds
(without alignments)
2649.143 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135915_26829/app_query.fasta_1.1628
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_8325_@runat_31122003_135915_26829 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1884	98.6	1546	9	AB011969	AB011969 Homo sapi
2	1849.5	96.8	1625	6	AR232183	AR232183 Sequence
3	1832	95.9	1272	6	AR086324	AR086324 Sequence
4	1832	95.9	1272	6	BD106653	BD106653 Method of
5	1832	95.9	1728	6	AX409750	AX409750 Sequence
6	1832	95.9	1728	6	AX706330	AX706330 Sequence
7	1832	95.9	1728	9	HUMPCPBX	M75106 Human prepr
8	1832	95.9	1749	6	I16100	I16100 Sequence 2
9	1832	95.9	1749	6	I33526	I33526 Sequence 2
10	1827	95.6	1272	9	BT006936	BT006936 Homo sapi
11	1827	95.6	1272	12	BT007990	BT007990 Synthetic
12	1827	95.6	1715	9	BC007057	BC007057 Homo sapi
13	1534.5	80.3	1421	10	AB021968	AB021968 Mus muscu
14	1534.5	80.3	1490	10	AF164524	AF164524 Mus muscu
15	1528.5	80.0	1430	10	AF186188	AF186188 Mus muscu
16	1502.5	78.6	1425	10	AB042598	AB042598 Rattus no
17	700.5	36.7	1327	5	AF190274	AF190274 Bothrops
18	682	35.7	1441	5	AB099302	AB099302 Paralicht
19	646	33.8	1251	4	SSC133775	AJ133775 Sus scrof
20	641	33.5	1254	12	BT008111	BT008111 Synthetic
21	641	33.5	1310	9	HSA224866	AJ224866 Homo sapi
22	641	33.5	1454	9	BC015338	BC015338 Homo sapi
23	640.5	33.5	1215	6	I67698	I67698 Sequence 1
24	639	33.4	1263	6	A51896	A51896 Sequence 60
25	639	33.4	1263	6	A87530	A87530 Sequence 11
26	639	33.4	1263	6	AR085861	AR085861 Sequence
27	631.5	33.0	1284	6	A51913	A51913 Sequence 77
28	631.5	33.0	1284	6	AR085875	AR085875 Sequence
29	628.5	32.9	1442	10	MUSCARA	J05118 Mouse mast
30	627	32.8	1251	4	DOGZAP47	D78348 Canis famil
31	625.5	32.7	1674	9	BC012613	BC012613 Homo sapi
32	623	32.6	2154	6	A67368	A67368 Sequence 12
33	621.5	32.5	1332	6	AX333659	AX333659 Sequence
34	621.5	32.5	1332	9	HUMPASP	M81057 Human proca
35	621	32.5	1427	10	RNU67914	U67914 Rattus norv
36	619.5	32.4	1622	9	HUMCARMC	M27717 Human mast
37	619.5	32.4	1622	11	G28614	G28614 human STS S
38	611.5	32.0	1302	6	AX616688	AX616688 Sequence
39	611.5	32.0	1311	6	AR225025	AR225025 Sequence
40	611.5	32.0	1311	6	AX106749	AX106749 Sequence
41	611.5	32.0	1993	6	AX704687	AX704687 Sequence
42	609.5	31.9	1907	9	BC033684	BC033684 Homo sapi
43	608.5	31.8	1672	5	BC041315	BC041315 Xenopus l
44	604	31.6	927	6	AR072911	AR072911 Sequence
45	581	30.4	1339	5	GGCBPAMR	X64539 G.domesticu

ALIGNMENTS

RESULT 1

```
AB011969
LOCUS       AB011969               1546 bp    mRNA    linear   PRI 02-FEB-2000
DEFINITION  Homo sapiens mRNA for carboxypeptidase B-like protein, complete
            cds.
ACCESSION   AB011969
VERSION     AB011969.1  GI:6855463
KEYWORDS    carboxypeptidase B-like protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)  1026?
  MEDLINE   92042093
   PUBMED   1939207
REFERENCE   2  (sites)
  AUTHORS   Matsumoto,A.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1546)
  AUTHORS   Matsumoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
            Medicine, Department of Radiation Biophysics & Genetics;
            Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:amat@med.kobe-u.ac.jp, Tel:81-78341-7451)
COMMENT     Sequence updated (10-Apr-1998).
FEATURES             Location/Qualifiers
     source          1. .1546
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             20. .1102
                     /note="carboxypeptidase B-like protein in human brain"
                     /codon_start=1
                     /product="carboxypeptidase B-like protein"
                     /protein_id="BAA90475.1"
                     /db_xref="GI:6855464"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQTGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDL
                     NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYS
                     QHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDD
                     WIYDLGIKYSFTSNPPVEKLLPLSLK"
     polyA_site      1546
                     /note="21 A nucleotides"
BASE COUNT      453 a    323 c    297 g    473 t
ORIGIN

Alignment Scores:
Pred. No.:              2.38e-176       Length:         1546
Score:                  1884.00         Matches:        354
Percent Similarity:     98.89%          Conservative:   2
Best Local Similarity:  98.33%          Mismatches:     4
Query Match:            98.59%          Indels:         0
DB:                     9               Gaps:           0

US-09-980-881A-2 (1-360) x AB011969 (1-1546)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGACTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 319

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||    ||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATT 559

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 619

Qy        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        620 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 679

Qy        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
              ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        680 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 739

Qy        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 799

Qy        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 859

Qy        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 919

Qy        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320
              ||||||||||||||||||:::|||    ||||||||||||||||||||||||||||||||
Db        920 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 979

Qy        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1039

Qy        341 IleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1099

RESULT 2
AR232183
LOCUS       AR232183               1625 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6455294.
ACCESSION   AR232183
VERSION     AR232183.1  GI:27274083
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
```

```
REFERENCE   1  (bases 1 to 1625)
  AUTHORS   Gan,W., Ketchum,K.A., Di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human protease proteins, nucleic acid molecules encoding
            human protease proteins, and uses thereof
  JOURNAL   Patent: US 6455294-A 1 24-SEP-2002;
FEATURES             Location/Qualifiers
     source          1. .1625
                     /organism="unknown"
BASE COUNT      488 a    334 c    315 g    488 t
ORIGIN

Alignment Scores:
Pred. No.:                6.52e-173      Length:          1625
Score:                    1849.50        Matches:         355
Percent Similarity:       94.18%         Conservative:    1
Best Local Similarity:    93.92%         Mismatches:      4
Query Match:              96.78%         Indels:          18
DB:                       6              Gaps:            1

US-09-980-881A-2 (1-360) x AR232183 (1-1625)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 316

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              |||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 556

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 616

Qy        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 676

Qy        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 736

Qy        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 796

Qy        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 856

Qy        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320
              ||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||||||
Db        917 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 976

Qy        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1036

Qy        341 IleLysTyrSerPhe---------------------------------------------- 345
              |||||||||||||||
Db       1037 ATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCG 1096

Qy        346 -------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                      ||||||||||||||||||||||||||||||||||||||||||||
Db       1097 GAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1148

RESULT 3
AR086324
LOCUS       AR086324               1272 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 1 from patent US 5985562.
ACCESSION   AR086324
VERSION     AR086324.1  GI:10013090
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.John. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk associated with plasma
            carboxypeptidase B polymorphisms
  JOURNAL   Patent: US 5985562-A 1 16-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="unknown"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN

Alignment Scores:
Pred. No.:                2.53e-171      Length:          1272
Score:                    1832.00        Matches:         357
Percent Similarity:       86.27%         Conservative:    1
Best Local Similarity:    86.02%         Mismatches:      2
Query Match:              95.87%         Indels:          55
DB:                       6              Gaps:            2

US-09-980-881A-2 (1-360) x AR086324 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        197 ------------------------------------------------------------ 197

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                     ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 4
BD106653
LOCUS       BD106653               1272 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Method of detecting thrombotic disease risk.
ACCESSION   BD106653
VERSION     BD106653.1  GI:23201471
KEYWORDS    JP 2002502253-A/1.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.J. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk
  JOURNAL   Patent: JP 2002502253-A 1 22-JAN-2002;
            SCHERING AG
COMMENT     PN   JP 2002502253-A/1
            PD   22-JAN-2002
            PF   02-JUN-1998 JP 1999501465
            PR   03-JUN-1997 US    08/869057
            PI   MICHAEL JOHN MORSER,MARIKO NAGASHIMA
            PC   C12Q1/68
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN

Alignment Scores:
Pred. No.:              2.53e-171       Length:         1272
Score:                  1832.00         Matches:        357
Percent Similarity:     86.27%          Conservative:   1
Best Local Similarity:  86.02%          Mismatches:     2
Query Match:            95.87%          Indels:         55
DB:                     6               Gaps:           2

US-09-980-881A-2 (1-360) x BD106653 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        197 ------------------------------------------------------------ 197

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                     ||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 5
AX409750
LOCUS       AX409750               1728 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2397 from Patent WO0229103.
ACCESSION   AX409750
VERSION     AX409750.1  GI:21442455
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2397 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M75106"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:              3.79e-171      Length:        1728
```

```
Score:                  1832.00       Matches:       357
Percent Similarity:     86.27%        Conservative:  1
Best Local Similarity:  86.02%        Mismatches:    2
Query Match:            95.87%        Indels:        55
DB:                     6             Gaps:          2

US-09-980-881A-2 (1-360) x AX409750 (1-1728)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        197 ------------------------------------------------------------ 197

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                     ||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
                ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy        344 SerPhe-------------------------------------------------------Th 346
                ||||||                                                       ||
Db       1160 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1219

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                |||||||||||||||||||||||||||||||||||||||||||
Db       1220 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 6
AX706330
LOCUS       AX706330               1728 bp    DNA     linear   PAT 04-APR-2003
DEFINITION  Sequence 1 from Patent WO03014391.
ACCESSION   AX706330
VERSION     AX706330.1  GI:29562754
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Morten,J.E. and Jackson,K.N.
  TITLE     Method
  JOURNAL   Patent: WO 03014391-A 1 20-FEB-2003;
            AstraZeneca AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:              3.79e-171      Length:         1728
Score:                  1832.00        Matches:        357
Percent Similarity:     86.27%         Conservative:   1
Best Local Similarity:  86.02%         Mismatches:     2
Query Match:            95.87%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-2 (1-360) x AX706330 (1-1728)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        197 ------------------------------------------------------------ 197

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy        344 SerPhe-------------------------------------------------------Th 346
              ||||||                                                       ||
Db       1160 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1219

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1220 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 7
HUMPCPBX
LOCUS       HUMPCPBX               1728 bp    mRNA    linear   PRI 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1  GI:189686
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
```

```
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 1728)
  AUTHORS  Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE    Isolation, molecular cloning, and partial characterization of a
           novel carboxypeptidase B from human plasma
  JOURNAL  J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE  92042093
   PUBMED  1939207
COMMENT    Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     gene            1. .1728
                     /gene="pCPB"
     CDS             20. .1291
                     /gene="pCPB"
                     /note="prepro-plasma carboxypeptidase B"
                     /codon_start=1
                     /protein_id="AAA60042.1"
                     /db_xref="GI:189687"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
     mat_peptide     86. .1288
                     /gene="pCPB"
                     /product="plasma carboxypeptidase B"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:                3.79e-171      Length:          1728
Score:                    1832.00        Matches:         357
Percent Similarity:       86.27%         Conservative:    1
Best Local Similarity:    86.02%         Mismatches:      2
Query Match:              95.87%         Indels:          55
DB:                       9              Gaps:            2

US-09-980-881A-2 (1-360) x HUMPCPBX (1-1728)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        197 ------------------------------------------------------------ 197

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1160 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1219

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1220 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 8
I16100
LOCUS       I16100                  1749 bp    DNA     linear   PAT 03-APR-1996
DEFINITION  Sequence 2 from patent US 5474901.
ACCESSION   I16100
VERSION     I16100.1  GI:1251008
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Antibodies to human carboxypeptidase B and methods of use thereof
  JOURNAL   Patent: US 5474901-A 2 12-DEC-1995;
```

```
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:              3.85e-171      Length:         1749
Score:                  1832.00        Matches:        357
Percent Similarity:     86.27%         Conservative:   1
Best Local Similarity:  86.02%         Mismatches:     2
Query Match:            95.87%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-2 (1-360) x I16100 (1-1749)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        197 ------------------------------------------------------------ 197

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              |||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 9
I33526
LOCUS       I33526                  1749 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 2 from patent US 5593674.
ACCESSION   I33526
VERSION     I33526.1  GI:1824317
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Plasma carboxypeptidase
  JOURNAL   Patent: US 5593674-A 2 14-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:              3.85e-171      Length:         1749
Score:                  1832.00        Matches:        357
Percent Similarity:     86.27%         Conservative:   1
Best Local Similarity:  86.02%         Mismatches:     2
Query Match:            95.87%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-2 (1-360) x I33526 (1-1749)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280
```

```
Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        197 ------------------------------------------------------------ 197

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                     ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 10
BT006936
LOCUS       BT006936                1272 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA,
            complete cds.
ACCESSION   BT006936
VERSION     BT006936.1  GI:30582710
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00156X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .1272
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma, carboxypeptidase
                     U)"
                     /protein_id="AAP35582.1"
                     /db_xref="GI:30582711"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      373 a    268 c    273 g    358 t
ORIGIN

Alignment Scores:
Pred. No.:              7.91e-171       Length:         1272
Score:                  1827.00         Matches:        356
Percent Similarity:     86.02%          Conservative:   1
Best Local Similarity:  85.78%          Mismatches:     3
Query Match:            95.60%          Indels:         55
DB:                     9               Gaps:           2

US-09-980-881A-2 (1-360) x BT006936 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
```

```
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60

Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80

Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100

Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120

Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140

Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160

Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180

Db        481 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197

Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        197 ------------------------------------------------------------ 197

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203

Db        661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223

Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243

Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263

Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283

Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303

Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323

Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343

Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        344 SerPhe------------------------------------------------------Th 346

Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360

Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243
```

```
RESULT 11
BT007990
LOCUS       BT007990                1272 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
            carboxypeptidase U) mRNA, partial cds.
ACCESSION   BT007990
VERSION     BT007990.1  GI:30584818
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00156L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .>1272
                     /note="Mutations: 1271:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens carboxypeptidase B2 (plasma,
                     carboxypeptidase U)"
                     /protein_id="AAP36662.1"
                     /db_xref="GI:30584819"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNVL"
BASE COUNT      372 a    268 c    273 g    359 t
ORIGIN

Alignment Scores:
Pred. No.:              7.91e-171       Length:         1272
Score:                  1827.00         Matches:        356
```

```
Percent Similarity:     86.02%          Conservative:  1
Best Local Similarity:  85.78%          Mismatches:    3
Query Match:            95.60%          Indels:        55
DB:                     12              Gaps:          2

US-09-980-881A-2 (1-360) x BT007990 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        197 ------------------------------------------------------------ 197

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        661 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              |||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              |||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        344 SerPhe----------------------------------------------------Th 346
              ||||||                                                    ||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 12
BC007057
LOCUS       BC007057               1715 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
            IMAGE:3934520, mRNA, complete cds.
ACCESSION   BC007057
VERSION     BC007057.1  GI:13937896
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1715)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 16 Row: f Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4503004.
FEATURES             Location/Qualifiers
     source          1. .1715
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:1361"
                     /db_xref="taxon:9606"
                     /clone="MGC:12495 IMAGE:3934520"
                     /tissue_type="Skeletal Muscle"
                     /clone_lib="NIH_MGC_81"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             25. .1296
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma)"
                     /protein_id="AAH07057.1"
                     /db_xref="GI:13937897"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
```

```
                    QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                    LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                    VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                    TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                    LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                    GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      518 a    348 c    341 g    508 t
ORIGIN

Alignment Scores:
Pred. No.:              1.17e-170       Length:         1715
Score:                  1827.00         Matches:        356
Percent Similarity:     86.02%          Conservative:   1
Best Local Similarity:  85.78%          Mismatches:     3
Query Match:            95.60%          Indels:         55
DB:                     9               Gaps:           2

US-09-980-881A-2 (1-360) x BC007057 (1-1715)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         25 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 84

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         85 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 144

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        145 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 204

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        205 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 264

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        265 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 324

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        325 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 384

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        385 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 444

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        445 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 504

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              |||||||||||||||||||||   |||||||||||||||||||||||||||||||||||||
Db        505 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 564

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              .|||||||||||||||||||||||||||||||||||||||||||||||||
Db        565 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 624

Qy        197 ------------------------------------------------------------ 197

Db        625 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 684

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                     ||||||||||||||||||
Db        685 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 744

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        745 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 804

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        805 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 864

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        865 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 924

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        925 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 984

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        985 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1044

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       1045 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1104

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1105 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1164

Qy        344 SerPhe-------------------------------------------------------Th 346
              ||||||                                                       ||
Db       1165 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1224

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1225 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1267

RESULT 13
AB021968
LOCUS       AB021968               1421 bp    mRNA    linear   ROD 27-JUL-2000
DEFINITION  Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION   AB021968
VERSION     AB021968.1  GI:9558447
KEYWORDS    carboxypeptidase R.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Obata,K., Okada,N.,
            Campbell,W. and Okada,H.
  TITLE     Pro-carboxypeptidase R is an acute phase protein in the mouse,
            whereas carboxypeptidase N is not
  JOURNAL   J. Immunol. 165 (2), 1053-1058 (2000)
  MEDLINE   20341711
   PUBMED   10878383
REFERENCE   2  (bases 1 to 1421)
  AUTHORS   Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1998) Tomoo Sato, Fukushimura Hospital, Choju
            Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
            441-8124, Japan (E-mail:tomoo@tcp-ip.or.jp, Tel:81-532-46-7511,
            Fax:81-532-46-4899)
FEATURES             Location/Qualifiers
     source          1. .1421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS             17. .1285
                     /codon_start=1
                     /product="carboxypeptidase R"
                     /protein_id="BAB03402.1"
                     /db_xref="GI:9558448"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
```

```
                FLLPERYIKPTCAEALAAISKIVWHVIRNT"
     polyA_site 1421
                /note="11 a nucleotides"
BASE COUNT      418 a    320 c    305 g    378 t
ORIGIN

Alignment Scores:
Pred. No.:              7.33e-142      Length:         1421
Score:                  1534.50        Matches:        302
Percent Similarity:     78.80%         Conservative:   25
Best Local Similarity:  72.77%         Mismatches:     32
Query Match:            80.30%         Indels:         56
DB:                     10             Gaps:           3

US-09-980-881A-2 (1-360) x AB021968 (1-1421)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||      |||   :::||||||   |||:::|||:::   |||||||||
Db         17 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTAT---GAGCAGCATGGC 73

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db         74 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 133

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db        134 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 193

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
                 |||   |||||||||:::|||||||||||||||||||||||||||||||||:::|||
Db        194 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 253

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
Db        254 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 313

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||:::||||||       ||||||||||||||||||||||||||||||||||||
Db        314 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 373

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              |||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db        374 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 433

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        434 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 493

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||      ||||||||||||||||||||||||||||||
Db        494 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 553

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        554 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 613

Qy        197 ------------------------------------------------------------ 197

Db        614 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 673

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        674 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 733

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        734 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 793

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        794 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 853

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        854 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 913

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              :::   ||||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        914 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 973

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        974 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1033

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::      :::||||||||||||||||||||||||||    |||||||||
Db       1034 GTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1093

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              :::|||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db       1094 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1153

Qy        344 SerPhe-------------------------------------------------Th 346
              ||||||                                                 ||
Db       1154 TCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATAC 1213

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||   |||:::|||||   |||   |||||||||
Db       1214 ATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1256

RESULT 14
AF164524
LOCUS       AF164524               1490 bp    mRNA    linear   ROD 05-APR-2000
DEFINITION  Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
            mRNA, complete cds.
ACCESSION   AF164524
VERSION     AF164524.1  GI:7416966
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1490)
  AUTHORS   Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van
            Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
  TITLE     Characterization of mouse thrombin-activatable fibrinolysis
            inhibitor
  JOURNAL   Thromb. Haemost. 83 (2), 297-303 (2000)
  MEDLINE   20201996
   PUBMED   10739389
REFERENCE   2  (bases 1 to 1490)
  AUTHORS   Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
            Utrecht 3584 CX, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .1490
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:477035"
     gene            1. .1490
                     /gene="Tafi"
     CDS             84. .1352
                     /gene="Tafi"
                     /function="carboxypeptidase"
                     /codon_start=1
                     /product="thrombin-activatable fibrinolysis inhibitor"
                     /protein_id="AAF62385.1"
                     /db_xref="GI:7416967"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
```

```
                    VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                    FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      438 a    337 c    322 g    393 t
ORIGIN

Alignment Scores:
Pred. No.:              7.8e-142      Length:         1490
Score:                  1534.50       Matches:        302
Percent Similarity:     78.80%        Conservative:   25
Best Local Similarity:  72.77%        Mismatches:     32
Query Match:            80.30%        Indels:         56
DB:                     10            Gaps:           3

US-09-980-881A-2 (1-360) x AF164524 (1-1490)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||      |||   :::||||||   |||:::|||:::   |||||||||
Db         84 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTAT---GAGCAGCATGGC 140

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db        141 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 200

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              :::||||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db        201 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 260

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
                 |||   |||||||||:::|||||||||||||||||||||||||||||||:::|||
Db        261 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 320

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||    ||||||   :::||||||:::   :::|||||||||
Db        321 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 380

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||:::||||||      ||||||||||||||||||||||||||||||||||||||
Db        381 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 440

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||    |||||||||:::|||
Db        441 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 500

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        501 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 560

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db        561 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 620

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        621 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 680

Qy        197 ------------------------------------------------------------ 197

Db        681 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 740

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||||
Db        741 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 800

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              |||||||||   :::   ||||||   |||:::||||||||||||   ||||||
Db        801 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 860

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        861 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 920

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              |||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        921 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 980

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              :::   ||||||||||||||||||||||||||||||||||   |||:::||||||||||
Db        981 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 1040

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1041 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1100

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::     :::||||||||||||||||||||||   |||||||||
Db       1101 GTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1160

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              :::|||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db       1161 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1220

Qy        344 SerPhe----------------------------------------------------Th 346
              ||||||                                                    ||
Db       1221 TCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATAC 1280

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||   |||:::||||||   |||   |||||||||
Db       1281 ATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1323
```

```
RESULT 15
AF186188
LOCUS       AF186188               1430 bp    mRNA    linear   ROD 01-OCT-1999
DEFINITION  Mus musculus carboxypeptidase U mRNA, complete cds.
ACCESSION   AF186188
VERSION     AF186188.1  GI:6003651
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Isolation and characterization of mouse liver carboxypeptidase B
            gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Division of Hematology, Washington
            University School of Medicine, 216 S. Kingshighway, St. Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="B6/CBAF1J"
                     /db_xref="taxon:10090"
     CDS             20. .1288
                     /note="pro-carboxypeptidase-B; thrombin-activatable
                     fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
                     plasminogen"
                     /codon_start=1
                     /product="carboxypeptidase U"
                     /protein_id="AAF00528.1"
                     /db_xref="GI:6003652"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVLTEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIGSFNKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      421 a    320 c    310 g    379 t
ORIGIN
```

Alignment Scores:

Pred. No.:	2.9e-141	Length:	1430
Score:	1528.50	Matches:	301
Percent Similarity:	78.55%	Conservative:	25
Best Local Similarity:	72.53%	Mismatches:	33
Query Match:	79.98%	Indels:	56
DB:	10	Gaps:	3

US-09-980-881A-2 (1-360) x AF186188 (1-1430)

```
Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              |||||||||      |||   :::||||||   |||:::|||:::   |||||||||
Db         20 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTAT---GAGCAGCATGGC 76

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db         77 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 136

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db        137 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 196

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
                 |||   |||||||||:::|||||||||||||||||||||||||||||||||:::|||
Db        197 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 256

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
Db        257 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 316

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||:::||||||       ||||||||||||||||||||||||||||||||||||||
Db        317 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 376

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||   :::|||||||:::|||
Db        377 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCTTAACTGAACAGCAT 436

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db        437 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 496

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 556

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db        557 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 616

Qy        197 ------------------------------------------------------------ 197

Db        617 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 676

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        677 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 736

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              |||||||||   :::   ||||||   |||:::|||||||||||||||   ||||||
Db        737 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 796

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db        797 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 856

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db        857 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 916

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              :::   ||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db        917 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 976

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1036

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||          :::||||||||||||||||||||||||   |||||||||
Db       1037 GTTCGTGCAATTGGAAGTTTTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1096

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              :::|||||||||||||||||||||   ||||||||||||||||||||||||||||||||
Db       1097 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1156

Qy        344 SerPhe----------------------------------------------------Th 346
              ||||||                                                    ||
Db       1157 TCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATAC 1216

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||   |||:::|||||   |||   |||||||||
Db       1217 ATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1259
```

Search completed: January 2, 2004, 00:15:59
Job time : 5594.33 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:00:56 ; Search time 62.7557 Seconds
(without alignments)
910.539 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1911	100.0	360	21	AAB11457	Human brain carbox
2	1822	95.3	386	24	ABG72095	Human protease pro
3	1817	95.1	386	24	ABG72093	Human protease pro
4	1799.5	94.2	423	17	AAR90293	Human plasma carbo
5	1799.5	94.2	423	18	AAW14733	Human plasma carbo
6	1799.5	94.2	423	20	AAW92270	Human plasma carbo
7	1794.5	93.9	423	14	AAR36273	Human plasma carbo
8	1785	93.4	338	21	AAB11458	Human brain carbox
9	1325	69.3	246	21	AAB11459	Human brain carbox
10	803.5	42.0	211	21	AAB58129	Lung cancer associ
11	641	33.5	417	24	ABB99688	Amino acid sequenc
12	640.5	33.5	404	16	AAR75131	Porcine Tyr-His-Me
13	639	33.4	415	17	AAW06172	Human pancreatic c
14	639	33.4	415	19	AAW74476	Human pancreatic c
15	631.5	33.0	424	17	AAW06175	ProHCPB with PelB
16	629.5	32.9	324	22	AAG66562	Bothops jararaca c
17	623	32.6	716	18	AAW41414	PreproHCPB-linker-
18	619.5	32.4	417	23	AAU84325	Protein CPA3 diffe
19	618.5	32.4	667	24	ABP97414	Anti-human seminal
20	604	31.6	307	17	AAW00602	Rat carboxypeptida
21	604	31.6	434	24	ABB99687	Amino acid sequenc
22	604	31.6	437	22	AAE00502	Human carboxypepti
23	604	31.6	437	23	AAE19175	Human protease, PR
24	576	30.1	306	16	AAR75132	Porcine carboxypep
25	576	30.1	306	22	AAU04477	Porcine carboxypep
26	563	29.5	354	22	AAG66560	Human secreted met
27	563	29.5	354	22	AAG66566	Human secreted met
28	563	29.5	362	22	AAB74682	Human protease and
29	563	29.5	374	22	AAG66547	Human secreted met
30	563	29.5	374	22	AAG66565	Human secreted met
31	563	29.5	374	23	AAU82703	Amino acid sequenc
32	563	29.5	374	24	ABJ37888	NOVX protein seque
33	563	29.5	374	24	ABG74658	Human zinc-binding
34	563	29.5	444	23	AAE24240	Human 23566 (carbo
35	561	29.4	329	17	AAW06174	Mature HCPB with P
36	561	29.4	349	17	AAW06173	mature HCPB-(His)6
37	554	29.0	349	17	AAW06181	Modified HCPB (D25
38	554	29.0	349	18	AAW13749	Carboxypeptidase B
39	553	28.9	349	17	AAW06182	Modified HCPB (D25
40	553	28.9	349	18	AAW13750	Carboxypeptidase B
41	550	28.8	349	18	AAW13763	Carboxypeptidase B
42	548	28.7	349	18	AAW13752	Carboxypeptidase B
43	548	28.7	349	18	AAW13758	Carboxypeptidase B
44	547	28.6	349	18	AAW13751	Carboxypeptidase B
45	547	28.6	374	24	ABJ37887	NOVX protein seque

ALIGNMENTS

RESULT 1
AAB11457
ID AAB11457 standard; Protein; 360 AA.
XX
AC AAB11457;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
DR N-PSDB; AAC81962.

```
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Claim 1; Page 68-71; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   360 AA;

  Query Match             100.0%;  Score 1911;  DB 21;  Length 360;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-188;
  Matches  360;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSET 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSET 240

Qy        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300

Qy        301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360


RESULT 2
ABG72095
ID   ABG72095 standard; Protein; 386 AA.
XX
AC   ABG72095;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #2.
XX
KW   Human; enzyme; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 169
FT                   /note= "This residue is the minor variant of a single
FT                         nucleotide polymorphism. The major variant is an
FT                         Ala residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Page -; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #2, the gene for which is located on chromosome 13.
CC   Note: This sequence is not shown in the specification but has been
CC   created by the indexer from information given in figure 3.
XX
SQ   Sequence   386 AA;

  Query Match             95.3%;  Score 1822;  DB 24;  Length 386;
  Best Local Similarity   98.8%;  Pred. No. 1.5e-178;
  Matches  342;  Conservative    1;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSET 240
              |||||||||||||||||||||||||||||||||| || ||||||||||||||||||||||
Db        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSET 240
```

```
Qy        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300

Qy        301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||:| |||||||||||||||||||||||||||||||||||||
Db        301 SEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346


RESULT 3
ABG72093
ID   ABG72093 standard; Protein; 386 AA.
XX
AC   ABG72093;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #1.
XX
KW   Human; enzyme; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 169
FT                   /note= "This residue is the major variant of a single
FT                         nucleotide polymorphism. The minor variant is an
FT                         Thr residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Fig 2; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #1, the gene for which is located on chromosome 13.
XX
SQ   Sequence   386 AA;

  Query Match             95.1%;  Score 1817;  DB 24;  Length 386;
  Best Local Similarity   98.6%;  Pred. No. 4.9e-178;
  Matches  341;  Conservative    1;  Mismatches    4;  Indels    0;  Gaps    0;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSET 240
              |||||||||||||||||||||||||||||||||||||| || ||||||||||||||||||
Db        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSET 240

Qy        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300

Qy        301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||:| |||||||||||||||||||||||||||||||||||||
Db        301 SEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346


RESULT 4
AAR90293
ID   AAR90293 standard; Protein; 423 AA.
XX
AC   AAR90293;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active-site     181
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    183
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    256..257
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     310
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    312
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
```

```
FT   Active-site     369
FT                   /note= "Residue suspecting of being involved in
FT                          catalytic activity of enzyme."
FT   Binding-site    370..371
FT                   /note= "Residues suspected of being part of the
FT                          binding site of enzyme."
FT   Binding-site    394
FT                   /note= "Residue suspected of being part of the
FT                          binding site of enzyme."
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   N-PSDB; AAT11671.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

  Query Match             94.2%;  Score 1799.5;  DB 17;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 3.6e-176;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              ||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 5
AAW14733
ID   AAW14733 standard; Protein; 423 AA.
XX
AC   AAW14733;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "Signal peptide"
FT   Protein         23..423
FT                   /note= "Mature PCPB"
FT   Cleavage-site   114
FT                   /note= "Potential clip site for activation of PCPB
FT                          as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    183
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     184
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    256
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    257
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     310
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    312
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     369
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    370
FT                   /note= "Expected to be involved in substrate
FT                          binding, thought to determine specificity
FT                          of PCPB as a carboxypeptidase B"
FT   Binding-site    371
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    394
FT                   /note= "Expected to be involved in substrate binding"
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   N-PSDB; AAT62846.
```

```
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence represents human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

  Query Match             94.2%;  Score 1799.5;  DB 18;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 3.6e-176;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              |||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 6
AAW92270
ID   AAW92270 standard; Protein; 423 AA.
XX
AC   AAW92270;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW   polymorphism detection; thrombotic disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "signal peptide"
FT   Protein         23..423
FT                   /note= "mature PCPBthr147"
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   N-PSDB; AAV74302.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Disclosure; Page 25-26; 35pp; English.
XX
CC   This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence   423 AA;

  Query Match             94.2%;  Score 1799.5;  DB 20;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 3.6e-176;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              |||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 7
AAR36273
ID   AAR36273 standard; Protein; 423 AA.
XX
AC   AAR36273;
```

```
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= signal_peptide
FT   Cleavage-site   114
FT                   /note= "putative site of cleavage by trypsin to
FT                          activate PCPB as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    183
FT                   /note= "expected to be involved in substrate
FT                          binding"
FT   Active-site     184
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    256..257
FT                   /note= "expected to be involved in substrate
FT                          binding"
FT   Active-site     310
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    312
FT                   /note= "expected to be involved in substrate
FT                          binding"
FT   Active-site     369
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    370..371
FT                   /note= "expected to be involved in substrate
FT                          binding; Asp 370 is the residue which
FT                          determines specificity of the PCPB as
FT                          a carboxypeptidase B"
FT   Binding-site    394
FT                   /note= "expected to be involved in substrate
FT                          binding"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   N-PSDB; AAQ41001.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Claim 2; Fig 4; 40pp; English.
XX
CC   This amino acid sequence was deduced from a clone isolated from a
CC   human liver cDNA library. There is about 40% sequence identity
CC   between the deduced preproPCPB amino acid sequence and pro-human
CC   mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC   carboxypeptidase B. Human PCPB has the same substrate binding sites
CC   as, and shares 6 cysteine residues which form 3 intramolecular
CC   disulphide bonds with, bovine and rat carboxypeptidase B. The
CC   presence of the same amino acid (Asp at position 348 of the mature
CC   protein) at the region in carboxypeptidases that determines
CC   substrate specificity, suggests that PCPB represents a plasma-
CC   derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC   tPA of plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence   423 AA;

  Query Match             93.9%;  Score 1794.5;  DB 14;  Length 423;
  Best Local Similarity   89.3%;  Pred. No. 1.2e-175;
  Matches  342;  Conservative    1;  Mismatches    3;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||||||||||| || |||||||| ||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGACSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              |||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 8
AAB11458
ID   AAB11458 standard; Protein; 338 AA.
XX
AC   AAB11458;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
```

```
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 71-73; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   338 AA;

  Query Match             93.4%;  Score 1785;  DB 21;  Length 338;
  Best Local Similarity   99.1%;  Pred. No. 7.8e-175;
  Matches  335;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60

Qy         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120

Qy        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 202
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180

Qy        203 KNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240

Qy        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322
              ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300

Qy        323 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360
              |||||||||||||||||||||||||||||||||||||
Db        301 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 338


RESULT 9
AAB11459
ID   AAB11459 standard; Protein; 246 AA.
XX
AC   AAB11459;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 74-75; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   246 AA;

  Query Match             69.3%;  Score 1325;  DB 21;  Length 246;
  Best Local Similarity   98.8%;  Pred. No. 1.1e-127;
  Matches  243;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60

Qy        175 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASS 234
              |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120

Qy        235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 294
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180

Qy        295 ELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 354
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240

Qy        355 LPLSLK 360
              ||||||
Db        241 LPLSLK 246


RESULT 10
AAB58129
ID   AAB58129 standard; Protein; 211 AA.
XX
AC   AAB58129;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
```

```
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18005.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 944-945; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   211 AA;

  Query Match             42.0%;  Score 803.5;  DB 21;  Length 211;
  Best Local Similarity   88.8%;  Pred. No. 4.6e-74;
  Matches  151;  Conservative    2;  Mismatches    8;  Indels    9;  Gaps    2;

Qy        185 WISPAFCLWFI--------GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSS 236
              |||   | | :         ||||||||||||||||||||| || ||||||||||||||
Db          3 WIS-MLCRWLMWMVMNYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 61

Qy        237 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 121

Qy        297 SLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              ||||||||||:|||||||||||||||||||||||||||||||||||||||
Db        122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 171


RESULT 11
ABB99688
ID   ABB99688 standard; Protein; 417 AA.
XX
AC   ABB99688;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Amino acid sequence of human zinc carboxypeptidase B precursor.
XX
KW   Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW   chronic obstructive pulmonary disease; COPD; haematological disorder;
KW   zinc carboxypeptidase B.
XX
OS   Homo sapiens.
XX
PN   WO200295018-A1.
XX
PD   28-NOV-2002.
XX
PF   16-MAY-2002; 2002WO-EP05392.
XX
PR   22-MAY-2001; 2001US-292329P.
XX
PA   (FARB ) BAYER AG.
XX
PI   Smolyar A;
XX
DR   WPI; 2003-120801/11.
XX
PT   New polynucleotide encoding a zinc carboxypeptidase B-like protein
PT   polypeptide, useful for treating diseases associated with
PT   carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT   haematological disorder  -
XX
PS   Disclosure; Fig 3; 113pp; English.
XX
CC   The present sequence represents a zinc carboxypeptidase B precursor.
CC   The specification describes a human zinc carboxypeptidase B-like
CC   protein. The zinc carboxypeptidase B-like protein is used for treating
CC   diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
CC   disease (COPD) or haematological disorders. The zinc carboxypeptidase
CC   B-like protein may also be used to identify compounds which may act
CC   as activators or inhibitors at the enzyme's active site, to raise
CC   specific antibodies which can block the enzyme and effectively reduce
CC   its activity, as a bait protein in a two-hybrid or three-hybrid assay
CC   to identify other proteins which bind to or interact with the human
CC   zinc carboxypeptidase-like protein polypeptide and modulate its
CC   activity, and for immunization of mammals.
XX
SQ   Sequence   417 AA;

  Query Match             33.5%;  Score 641;  DB 24;  Length 417;
  Best Local Similarity   35.3%;  Pred. No. 7.7e-57;
  Matches  135;  Conservative   71;  Mismatches  132;  Indels   44;  Gaps    7;

Qy          3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADL 61
              : :| ||| : |     |    |: :|          : ::: | :| :|  |:| :
Db          1 MLALLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQ 60

Qy         62 IVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYE 121
              |      | | | | |   |:  |  : :   ||:::: :::: |  :     ||:   ||
Db         61 IKPHSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYE 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||::||| |
Db        118 KYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFH 176

Qy        182 AREWISPAFCLWFI----------------------------GH------NRMWRKN 204
              |||||||||| ||:                            |:      :| |||
Db        177 AREWISPAFCQWFVREAVRTYGREIQVTELLNKLDFYVLPVLNIDGYIYTWTKSRFWRKT 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              || :  : |||||  | ||    ||| ||| : | |||||   ||| | ||:| |:|  ::
Db        237 RSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLS 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              ||||:::||||| :::|||||     ::: ||: :|   |: :  :   |:||:| |: |
Db        296 SIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATT 354

Qy        325 LYLAPGGGDDWIYDLGIKYSFT 346
              :| | || ||| || ||:||||
Db        355 IYPAAGGSDDWAYDQGIRYSFT 376


RESULT 12
AAR75131
ID   AAR75131 standard; Protein; 404 AA.
XX
AC   AAR75131;
XX
DT   13-MAR-1996  (first entry)
```

```
XX
DE   Porcine Tyr-His-Met Procarboxypeptidase B.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB.
XX
OS   Sus scrofa.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   N-PSDB; AAQ90600.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Example 1; Page 16-17; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence   404 AA;

  Query Match             33.5%;  Score 640.5;  DB 16;  Length 404;
  Best Local Similarity   37.7%;  Pred. No. 8.3e-57;
  Matches  136;  Conservative   60;  Mismatches  122;  Indels   43;  Gaps    6;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
              |:  :|          : :|  | :| :|  |:| :   |     | | | | |:  |:
Db         10 FEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 69

Qy         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
               |  : :   ||: ::  ::: |  :     | :   ||:|::   | :| : :|  :||
Db         70 FLEQNELQYEVLINNLRSVLEAQFDSRV---RTTGHSYEKYNNWETIEAWTKQVTSENPD 126

Qy        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
              ::::  ||::|    :|:||| ||    | ||::||| ||||||| ||| ||:
Db        127 LISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCGFHAREWISHAFCQWFVREAVLTY 185

Qy        196 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                   |:      ||||| ||  |   ||||| | ||
Db        186 GYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGTTCIGTDPNRNF-DA 244

Qy        226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
               ||  |||:  | |||||   ||| | ||:| |:| |:: ||||:::|||| |::||||
Db        245 GWCTTGASTDPCDETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 304

Qy        286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
                :::  ||: :|  ||: :  |   |:||:| |: |:| | || ||| || ||||||
Db        305 DYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGATTIYPAAGGSDDWAYDQGIKYSF 363

Qy        346 T 346
              |
Db        364 T 364
```

```
RESULT 13
AAW06172
ID   AAW06172 standard; Protein; 415 AA.
XX
AC   AAW06172;
XX
DT   12-FEB-1997  (first entry)
XX
DE   Human pancreatic carboxypeptidase B.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..13
FT                   /note= "partial enzyme secretion leader sequence;
FT                        (pre-sequence)"
FT   Peptide         14..108
FT                   /label= pro-sequence
FT   Protein         109..415
FT                   /label= mature_enzyme
FT   Misc-difference 243
FT                   /note= "not observed in other published HPCB
FT                         sequences"
FT   Active-site     361
FT                   /note= "Asp253 characteristic of B type specificity"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   N-PSDB; AAT42494.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 14; Page 131-132; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is that of the native prepro-HCPB.
XX
SQ   Sequence   415 AA;

  Query Match             33.4%;  Score 639;  DB 17;  Length 415;
  Best Local Similarity   35.6%;  Pred. No. 1.2e-56;
  Matches  135;  Conservative   69;  Mismatches  131;  Indels   44;  Gaps    7;
```

```
Qy          6 LAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK 64
              | ||| : |    |    |:  :|          : ::: | :| :|  |:| :   |
Db          2 LLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKP 61

Qy         65 KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYH 124
               | | | | |   |:  |  : :   ||:::: :::: |  :     ||:   ||:|:
Db         62 HSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYN 118

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
                 | :| : :   :| :::: ||::||   :|:||| ||    | ||::||| ||||
Db        119 KWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHARE 177

Qy        185 WISPAFCLWFI------------------------------GH------NRMWRKNRSF 207
              ||||||| ||:                             |:      :| ||| ||
Db        178 WISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRST 237

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              :  : ||||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  :: ||
Db        238 HTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIK 296

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              ||:::||||| :::|||     ::: ||: :|   |: :  :   |:||:| |: |:|
Db        297 AYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYP 355

Qy        328 APGGGDDWIYDLGIKYSFT 346
              | || ||| || ||:||||
Db        356 AAGGSDDWAYDQGIRYSFT 374
```

RESULT 14
AAW74476
ID AAW74476 standard; Protein; 415 AA.
XX
AC AAW74476;
XX
DT 20-NOV-1998 (first entry)
XX
DE Human pancreatic carboxypeptidase protein sequence.
XX
KW Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW prodrug therapy.
XX
OS Homo sapiens.
XX
PN WO9835988-A1.
XX
PD 20-AUG-1998.
XX
PF 10-FEB-1998; 98WO-GB00415.
XX
PR 29-OCT-1997; 97GB-0022727.
PR 14-FEB-1997; 97GB-0003104.
PR 18-OCT-1997; 97GB-0022003.
XX
PA (ZENE) ZENEPA (ZENE) ZENECA LTD.
XX
PI Edge MD;
XX
DR WPI; 1998-467168/40.
DR N-PSDB; AAV41795.
XX
PT New modified pro-domain of carboxy-peptidase B - enhances expression
PT of co-expressed proteins for production of recombinant
PT carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT enzyme prodrug therapy
XX
PS Example 1; Page 54-55; 83pp; English.
XX
CC The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC modified pro-domain of CPB on a separate gene to enhance recombinant
CC expression. This process can be used to produce recombinant CPB in
CC eukaryotic cells, or fusions of CPB with antibody chains. CPB is used in
CC insulin production and protein sequencing, while its fusions with
CC antibodies are useful in antibody-directed enzyme prodrug therapy. The
CC modified pro-domain provide increased yields of recombinant CPB, possibly
CC by protecting the C-terminus against enzymatic degradation or by
CC increasing intracellular trafficking.
XX
SQ Sequence 415 AA;

Query Match 33.4%; Score 639; DB 19; Length 415;
Best Local Similarity 35.6%; Pred. No. 1.2e-56;
Matches 135; Conservative 69; Mismatches 131; Indels 44; Gaps 7;

```
Qy          6 LAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK 64
              | ||| : |    |    |:  :|          : ::: | :| :|  |:| :   |
Db          2 LLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKP 61

Qy         65 KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYH 124
               | | | | |   |:  |  : :   ||:::: :::: |  :     ||:   ||:|:
Db         62 HSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYN 118

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
                 | :| : :   :| :::: ||::||   :|:||| ||    | ||::||| ||||
Db        119 KWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHARE 177

Qy        185 WISPAFCLWFI------------------------------GH------NRMWRKNRSF 207
              ||||||| ||:                             |:      :| ||| ||
Db        178 WISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRST 237

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              :  : ||||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  :: ||
Db        238 HTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIK 296

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              ||:::||||| :::|||     ::: ||: :|   |: :  :   |:||:| |: |:|
Db        297 AYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYP 355

Qy        328 APGGGDDWIYDLGIKYSFT 346
              | || ||| || ||:||||
Db        356 AAGGSDDWAYDQGIRYSFT 374
```

RESULT 15
AAW06175
ID AAW06175 standard; Protein; 424 AA.
XX
AC AAW06175;
XX
DT 17-FEB-1997 (first entry)
XX
DE ProHCPB with PelB leader sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= PelB_leader_sequence
FT Peptide 23..117
FT /label= pro_sequence
FT Protein 118..424
FT /label= mature_HCPB
XX
PN WO9620011-A1.
XX
PD 04-JUL-1996.
XX
PF 21-DEC-1995; 95WO-GB02991.
XX

```
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   N-PSDB; AAT42506.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 18; Page 140-142; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is that of proHCPB with a PelB leader
CC   sequence encoded by pICI1738.
XX
SQ   Sequence   424 AA;

  Query Match             33.0%;  Score 631.5;  DB 17;  Length 424;
  Best Local Similarity   35.7%;  Pred. No. 7.6e-56;
  Matches  129;  Conservative   68;  Mismatches  121;  Indels   43;  Gaps    6;

Qy          23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
               |:  :|         : ::: | :| :|  |:| :   |     | | | | |   |:
Db          29 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 88

Qy          83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
                 |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db          89 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 145

Qy         143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
               ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db         146 LISRSVIGTTPEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 204

Qy         196 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                    |:      :| ||| || :  : ||||| | ||
Db         205 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 263

Qy         226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
               ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||||  :::|||| 
Db         264 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 323

Qy         286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
                 ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db         324 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 382

Qy         346 T 346
               |
Db         383 T 383


Search completed: January  1, 2004, 19:28:05
Job time : 64.7557 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 10:51:23 ; Search time 71.6472 Seconds
(without alignments)
6653.349 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1025.2	94.9	1625	4	US-09-813-133A-1	Sequence 1, Appli
2	909	84.2	1272	2	US-08-869-057-1	Sequence 1, Appli
3	909	84.2	1749	1	US-07-649-591B-2	Sequence 2, Appli
4	909	84.2	1749	1	US-08-277-540-2	Sequence 2, Appli
5	909	84.2	1749	1	US-08-430-787A-2	Sequence 2, Appli
6	204	18.9	55827	4	US-09-813-133A-3	Sequence 3, Appli
7	124	11.5	927	2	US-08-782-760-5	Sequence 5, Appli
8	124	11.5	927	5	PCT-US96-00995-5	Sequence 5, Appli
9	123.4	11.4	921	1	US-08-696-139-3	Sequence 3, Appli
10	123.4	11.4	1215	1	US-08-696-139-1	Sequence 1, Appli
11	109	10.1	1263	4	US-09-011-769A-38	Sequence 38, Appl
12	107.4	9.9	999	2	US-08-860-882A-67	Sequence 67, Appl
13	107.4	9.9	999	4	US-09-011-769A-50	Sequence 50, Appl
14	107.4	9.9	1053	2	US-08-860-882A-64	Sequence 64, Appl
15	107.4	9.9	1053	4	US-09-011-769A-46	Sequence 46, Appl
16	107.4	9.9	1263	2	US-08-860-882A-56	Sequence 56, Appl
17	107.4	9.9	1284	2	US-08-860-882A-71	Sequence 71, Appl
18	107.4	9.9	1284	4	US-09-011-769A-55	Sequence 55, Appl
19	104.2	9.6	1053	4	US-09-463-451-27	Sequence 27, Appl
c 20	104.2	9.6	1053	4	US-09-463-451-28	Sequence 28, Appl
21	104.2	9.6	1059	2	US-08-860-882A-74	Sequence 74, Appl
22	104.2	9.6	1059	2	US-08-860-882A-77	Sequence 77, Appl
23	104.2	9.6	1059	4	US-09-011-769A-59	Sequence 59, Appl
24	104.2	9.6	1059	4	US-09-011-769A-63	Sequence 63, Appl
25	101	9.4	1870	3	US-09-171-945-112	Sequence 112, App
26	101	9.4	2154	3	US-09-171-945-124	Sequence 124, App
27	92.8	8.6	1311	4	US-09-675-305-9	Sequence 9, Appli
28	92.8	8.6	2128	4	US-09-675-305-13	Sequence 13, Appl
29	88.6	8.2	1050	4	US-09-675-305-11	Sequence 11, Appl
30	87.6	8.1	1200	4	US-09-710-099-7	Sequence 7, Appli
31	83.6	7.7	945	4	US-09-710-099-3	Sequence 3, Appli
32	83.6	7.7	945	4	US-09-710-099-11	Sequence 11, Appl
33	78.4	7.3	741	4	US-09-675-305-5	Sequence 5, Appli
34	71.6	6.6	1251	3	US-08-640-906-3	Sequence 3, Appli
35	71.6	6.6	1251	4	US-09-395-936-3	Sequence 3, Appli
36	70.6	6.5	1257	3	US-08-640-906-1	Sequence 1, Appli
37	70.6	6.5	1257	4	US-09-395-936-1	Sequence 1, Appli
38	51.4	4.8	1056	4	US-09-710-099-1	Sequence 1, Appli
39	51.4	4.8	1056	4	US-09-710-099-9	Sequence 9, Appli
40	51.4	4.8	1311	4	US-09-710-099-5	Sequence 5, Appli
41	51.4	4.8	2247	4	US-09-710-099-15	Sequence 15, Appl
42	48	4.4	629	4	US-09-280-116-228	Sequence 228, App
43	46.8	4.3	285	2	US-08-782-760-2	Sequence 2, Appli
44	46.8	4.3	285	5	PCT-US96-00995-2	Sequence 2, Appli
45	40.8	3.8	10411	4	US-08-961-527-89	Sequence 89, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Query Match 94.9%; Score 1025.2; DB 4; Length 1625;
Best Local Similarity 99.2%; Pred. No. 7.4e-306;
Matches 1030; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 316

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 556

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 616

Qy        601 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 676

Qy        661 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 720
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 736

Qy        721 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 796

Qy        781 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 856

Qy        841 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy        901 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 960
              |||||||||||||||||||||| |||| ||||||||||||||||||||||||||||||||
Db        917 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 976

Qy        961 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1036

Qy       1021 ATCAAATATTCGTTTACA 1038
              ||||||||||||||||||
Db       1037 ATCAAATATTCGTTTACA 1054


RESULT 2
US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1272 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    PUBLICATION INFORMATION:
;      AUTHORS:  Eaton, Dan L.
;      AUTHORS:  Malloy, Beth E.
;      AUTHORS:  Tsai, Siao P
;      AUTHORS:  Henzel, William
;      AUTHORS:  Drayna, Dennis
;      TITLE:  Isolation, Molecular Cloning, and Partial
;      TITLE:  Characterization of a No. 5985562el Carboxypeptidase B
;      TITLE:  from Human Plasma
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  266
;      ISSUE:  32
;      PAGES:  21833-21838
;      DATE:  No. 5985562 15-1991
US-08-869-057-1

  Query Match           84.2%;  Score 909;  DB 2;  Length 1272;
  Best Local Similarity  89.9%;  Pred. No. 4.2e-270;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600
```

```
Qy        592 ------------------------------------------------------------ 591

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        592 -------------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1141 TCGTTTACA 1149
```

```
RESULT 3
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  hybridization probe
;      LOCATION:  133 to 178
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  potential clip site
;      LOCATION:  380 to 382
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  signal sequence
;      LOCATION:  41 to 106
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-07-649-591B-2

  Query Match             84.2%;  Score 909;  DB 1;  Length 1749;
  Best Local Similarity   89.9%;  Pred. No. 4.8e-270;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591
```

```
Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                      ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 4
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-277-540-2

  Query Match            84.2%;  Score 909;  DB 1;  Length 1749;
  Best Local Similarity  89.9%;  Pred. No. 4.8e-270;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ---------------------------------------------------------- 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                      ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-787A-2

  Query Match            84.2%;  Score 909;  DB 1;  Length 1749;
  Best Local Similarity  89.9%;  Pred. No. 4.8e-270;
  Matches 1033;  Conservative    0;  Mismatches    5;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        592 ------------------------------------------------------------ 591

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        592 -----------------------------------------AATCGAATGTGGAGAAAG 609
                                                        ||||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000
```

```
Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1181 TCGTTTACA 1189


RESULT 6
US-09-813-133A-3
; Sequence 3, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001173
;  CURRENT APPLICATION NUMBER: US/09/813,133A
;  CURRENT FILING DATE:  2001-06-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 55827
;   TYPE: DNA
;   ORGANISM: Human
US-09-813-133A-3

  Query Match             18.9%;  Score 204;  DB 4;  Length 55827;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-52;
  Matches  204;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48643 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 48702

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48703 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 48762

Qy        805 TACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGT 864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      48763 TACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGT 48822

Qy        865 AAAAGCAAAGACCATGAGGAACTG 888
              ||||||||||||||||||||||||
Db      48823 AAAAGCAAAGACCATGAGGAACTG 48846


RESULT 7
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
;  GENERAL INFORMATION:
;    APPLICANT:  Hartman, Jacob
;    APPLICANT:  Fulga, Netta
;    APPLICANT:  Mendelovitch, Simona
;    APPLICANT:  Gorecki, Marian
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,760
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/378,233
;      FILING DATE:  25-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
US-08-782-760-5

  Query Match             11.5%;  Score 124;  DB 2;  Length 927;
  Best Local Similarity   57.9%;  Pred. No. 2.6e-28;
  Matches  278;  Conservative    0;  Mismatches  195;  Indels    7;  Gaps    3;

Qy        596 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 655
              |||||||||||||| | ||| |||       ||     | |   ||| | ||   |||||  |
Db        362 GAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGACCCCA 421

Qy        656 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 715
              | || || |||     |       |||||||||  |  || ||  ||     | |||||  |
Db        422 ACAGGAATTTTAATGC---TGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478

Qy        716 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 775
              ||||  ||||||||||     || ||||| |||  |||    |||||  ||||   ||||
Db        479 AAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGATTTCA 538

Qy        776 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 835
              |  | |  ||  ||  |   || || || ||| | | ||| || ||||||||| |||   |
Db        539 TCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAGATGA 598

Qy        836 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 895
              |  | |  || || ||||||       |         | || |||||||| ||  |
Db        599 TGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAATGCCC 658

Qy        896 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 955
              | |  |  || |     |  ||| |  |||    |    |||| ||| ||| |||
Db        659 TGGTGAAAGGTGCGGCAAAGGAGCTTGCC---ACTCTGCATGGCACCAAGTACACATATG 715

Qy        956 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1015
              |||  ||  |  |||  | || |  ||| |||| ||     ||||| |||   ||||||
Db        716 GCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTATGATC 775

Qy       1016 TGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTC 1075
                |||  |||||||||| ||||| |   |  | || | | | | |||| |   ||| ||||
Db        776 AGGGAATCAAATATTCCTTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834
```

```
RESULT 8
PCT-US96-00995-5
; Sequence 5, Application PC/TUS9600995
;  GENERAL INFORMATION:
;    APPLICANT:  Bio-Technology General Corp.
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00995
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847-A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
PCT-US96-00995-5

  Query Match            11.5%;  Score 124;  DB 5;  Length 927;
  Best Local Similarity  57.9%;  Pred. No. 2.6e-28;
  Matches  278;  Conservative    0;  Mismatches  195;  Indels    7;  Gaps    3;

Qy        596 GAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGA 655
              |||||||||||||  | ||| |||       ||    | |   ||| | ||   |||||  |
Db        362 GAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGACCCCA 421

Qy        656 ATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGG 715
              | || || |||    |     |||||||| |  || || ||  |     | |||||  |
Db        422 ACAGGAATTTTAATGC---TGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGCTCTG 478

Qy        716 AAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCT 775
              ||||  ||||||||||      || ||||| |||  |||     |||||  ||||   ||||
Db        479 AAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGATTTCA 538

Qy        776 TGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATA 835
              |  |  ||  ||  |   || || || ||| | | ||| || ||||||||  |||   |
Db        539 TCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAGATGA 598

Qy        836 TAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAG 895
              |  |  ||  || ||||||      |         | || ||||||| ||  |
Db        599 TGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAATGCCC 658

Qy        896 TAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATG 955
              |  |  |  || |     |  ||| |  |||     |    |||| ||| ||| |||
Db        659 TGGTGAAAGGTGCGGCAAAGGAGCTTGCC---ACTCTGCATGGCACCAAGTACACATATG 715

Qy        956 GCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATT 1015
              |||  ||  |  |||  | || |  ||| |||| ||    ||||| |||   ||||||
Db        716 GCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTATGATC 775

Qy       1016 TGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTC 1075
              ||| |||||||||||| ||||| |   |  | || | | | |||| |   ||| ||||
Db        776 AGGGAATCAAATATTCCTTTACCTTTGA-ACTCCGGGATACAGGCTTCTTTGGCTTTCTC 834


RESULT 9
US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..921
US-08-696-139-3

  Query Match            11.4%;  Score 123.4;  DB 1;  Length 921;
  Best Local Similarity  56.3%;  Pred. No. 3.9e-28;
  Matches  273;  Conservative    0;  Mismatches  206;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |     |    | | || ||||||||||||||| |||
Db        321 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCG 380

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              |||  | ||||    | |   ||||| || |||||||  || || || ||||    |
Db        381 CTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTTGATGCTGG 440
```

```
Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              |||||       || || || |    | | |||   || || |||||||||  |
Db        441 G---TGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGC 497

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | | ||||| |||  |||     |||||  |||||  |||  |  | |  ||  ||  |
Db        498 TGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTC 557

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || |||||||||  | |   || || ||||||||  |||   ||| | |  || ||||||||
Db        558 CATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTA 617

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              |     |     |  ||  ||| ||  ||  |   | || |  |  || ||
Db        618 TGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAA 677

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |   |||     ||     |    |||| ||| ||| | ||||  ||  |   |||  |
Db        678 AGAACTTG---CTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAAT 734

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    || || |||   |||||    || ||||||||||| ||
Db        735 CTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTT 794

Qy       1035 TACAT 1039
                |||
Db        795 CACCT 799


RESULT 10
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1215 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1215
US-08-696-139-1

  Query Match           11.4%;  Score 123.4;  DB 1;  Length 1215;
  Best Local Similarity   56.3%;  Pred. No. 4.5e-28;
  Matches  273;  Conservative    0;  Mismatches  206;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |     |   | | || ||||||||||||||||  |||
Db        615 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCG 674

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              |||  | ||||    | |   ||||| || |||||||  || || || ||||    |
Db        675 CTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTTGATGCTGG 734

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              |||||         || || || |    | | |||   || || |||||||||  |
Db        735 G---TGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGC 791

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | | ||||| |||  |||     |||||  |||||  |||  |  | |  ||  ||  |
Db        792 TGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTC 851

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || |||||||||  | |   || || ||||||||  |||   ||| | |  || ||||||||
Db        852 CATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTA 911

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              |     |     |  ||  ||| ||  ||  |   | || |  |  || ||
Db        912 TGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAA 971

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |   |||     ||     |    |||| ||| ||| | ||||  ||  |   |||  |
Db        972 AGAACTTG---CTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAAT 1028

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    || || |||   |||||    || ||||||||||| ||
Db       1029 CTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTT 1088

Qy       1035 TACAT 1039
                |||
Db       1089 CACCT 1093


RESULT 11
US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
```

```
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 38:
;       SEQUENCE CHARACTERISTICS:
;             LENGTH: 1263 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;       SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

  Query Match             10.1%;  Score 109;  DB 4;  Length 1263;
  Best Local Similarity   54.4%;  Pred. No. 1.3e-23;
  Matches  264;  Conservative    0;  Mismatches  215;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |    |     | | |  ||| | ||||||||||  ||
Db        645 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 704

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              ||   | || |      |   ||||| || |||||||  || || || ||||   |
Db        705 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 761

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              ||||||||    || || ||  |   | | ||    ||||| ||||||||||  |
Db        762 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 821

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | ||||| |||   |||   |||||  |||||  |||| |  | |  ||  ||
Db        822 CGCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 881

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        882 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 941

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              | |    |      |  |  || |||  ||  ||  |    | || |  |   | ||
Db        942 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 1001

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |  ||| |    |    |   |||| ||| ||| ||||||  ||  |  |||  |
Db       1002 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 1058

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    ||||| |||   |||||    || |||| ||||||  ||
Db       1059 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 1118

Qy       1035 TACAT 1039
              ||| |
Db       1119 CACCT 1123


RESULT 12
US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  67:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  999 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-67

  Query Match             9.9%;  Score 107.4;  DB 2;  Length 999;
  Best Local Similarity   54.2%;  Pred. No. 3.5e-23;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |    |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              ||   | || |      |   ||||| || |||||||  || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              ||||||||    || || ||  |   | | ||    ||||| ||||||||||  |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | |||||  |||   |||||  |||||  |||| |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              | |    |      |  |  || |||  ||  ||  |    | || |  |   | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |  ||| |    |    |   |||| ||| ||| ||||||  ||  |  |||  |
```

```
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    ||||| |||   |||||    || |||| |||||| ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1035 TACAT 1039
              ||| |
Db        861 CACCT 865


RESULT 13
US-09-011-769A-50
; Sequence 50, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 50:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 999 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION: 1..987
;        FEATURE:
;             NAME/KEY:  mat_peptide
;             LOCATION: 67..987
;        SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-011-769A-50

  Query Match             9.9%;  Score 107.4;  DB 4;  Length 999;
  Best Local Similarity   54.2%;  Pred. No. 3.5e-23;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |    |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              ||   | || |      |   |||| || |||||||  || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              ||||||||    || || ||  |   | | ||    ||||| |||||||||| | |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | ||||| |||   ||    |||||  |||||  ||||  |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              | |    |     |  |  || |||  ||  ||  |    | || |  |   | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |   ||| |    |      |    |||| ||| ||| ||||||  ||  |   ||| |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    ||||| |||   |||||    || |||| |||||| ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1035 TACAT 1039
              ||| |
Db        861 CACCT 865


RESULT 14
US-08-860-882A-64
; Sequence 64, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
```

```
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1053 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-64

  Query Match             9.9%;  Score 107.4;  DB 2;  Length 1053;
  Best Local Similarity   54.2%;  Pred. No. 3.6e-23;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |     |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              ||   | || |      |   ||||| || ||||||| || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              ||||||||    || || ||  |   | | ||    ||||| |||||||||| |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | ||||| |||   ||     |||||  |||||  |||| |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              | |    |       |  |  || |||  ||  ||  |    | || |  |    | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |   ||| |    |      |    |||| ||| ||| ||||||  ||  |   ||| |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    ||||| |||   |||||    || |||| ||||||  ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1035 TACAT 1039
              ||| |
Db        861 CACCT 865


RESULT 15
US-09-011-769A-46
; Sequence 46, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1053 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION: 1..1047
;       FEATURE:
;            NAME/KEY:  mat_peptide
;            LOCATION: 67..1047
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-011-769A-46

  Query Match             9.9%;  Score 107.4;  DB 4;  Length 1053;
  Best Local Similarity   54.2%;  Pred. No. 3.6e-23;
  Matches  263;  Conservative    0;  Mismatches  216;  Indels    6;  Gaps    2;

Qy        555 GATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGAAAGAACCG 614
              | ||  |  || |  || | | |     |     | | |  ||| | ||||||||||  ||
Db        387 GCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCG 446

Qy        615 TTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTCTCCAA 674
              ||   | || |      |   ||||| || ||||||| || || || ||||   |
Db        447 CTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTTGATGC--- 503

Qy        675 ACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTA 734
              ||||||||    || || ||  |   | | ||    ||||| |||||||||| |
Db        504 TGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGC 563

Qy        735 TCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCA 794
              | ||||| |||   ||     |||||  |||||  |||| |  | |  ||  ||
Db        564 CGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTC 623

Qy        795 GATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTA 854
              || || |||||  | |  || || || ||||||||    ||  | |  || || || ||
Db        624 CATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATA 683

Qy        855 TACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCG 914
              | |    |       |  |  || |||  ||  ||  |    | || |  |    | ||
Db        684 TGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAA 743

Qy        915 TGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTT 974
              |   ||| |    |      |    |||| ||| ||| ||||||  ||  |   ||| |
Db        744 AGAACTTGCC---TCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAAT 800

Qy        975 ATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTT 1034
              || |  ||| |||| ||    ||||| |||   |||||    || |||| ||||||  ||
Db        801 CTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTT 860

Qy       1035 TACAT 1039
              ||| |
Db        861 CACCT 865
```

Search completed: January 1, 2004, 15:41:02
Job time : 78.6472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 13:56:00 ; Search time 372.077 Seconds
(without alignments)
10046.789 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	909	84.2	1728	10	US-09-880-107-2396	Sequence 2396, Ap
	2	854.6	79.1	1272	13	US-10-379-836-1	Sequence 1, Appli
	3	441	40.8	1400	9	US-09-925-302-24	Sequence 24, Appl
	4	331.2	30.7	416	10	US-09-960-352-14595	Sequence 14595, A
	5	282.4	26.1	397	10	US-09-960-352-8425	Sequence 8425, Ap
	6	131	12.1	1547	13	US-10-027-632-265133	Sequence 265133,
	7	131	12.1	1547	13	US-10-027-632-265134	Sequence 265134,
	8	131	12.1	1547	13	US-10-027-632-265135	Sequence 265135,
	9	131	12.1	1547	13	US-10-027-632-265136	Sequence 265136,
	10	131	12.1	1547	14	US-10-027-632-265133	Sequence 265133,
	11	131	12.1	1547	14	US-10-027-632-265134	Sequence 265134,
	12	131	12.1	1547	14	US-10-027-632-265135	Sequence 265135,
	13	131	12.1	1547	14	US-10-027-632-265136	Sequence 265136,
c	14	129.6	12.0	431	10	US-09-917-800A-468	Sequence 468, App
	15	107.6	10.0	1254	15	US-10-229-546-3	Sequence 3, Appli
	16	107.6	10.0	1622	12	US-10-429-802-21	Sequence 21, Appl
	17	107.6	10.0	1622	15	US-10-229-546-1	Sequence 1, Appli
	18	107.6	10.0	1633	13	US-10-341-434-187	Sequence 187, App
	19	101	9.4	1870	10	US-09-910-059-112	Sequence 112, App
	20	101	9.4	2154	10	US-09-910-059-124	Sequence 124, App
	21	94.6	8.8	1332	10	US-09-954-456-1141	Sequence 1141, Ap
	22	92.8	8.6	1311	14	US-10-200-344-9	Sequence 9, Appli
	23	92.8	8.6	1993	12	US-10-274-639-33	Sequence 33, Appl
	24	92.8	8.6	2128	14	US-10-200-344-13	Sequence 13, Appl
	25	91.6	8.5	936	15	US-10-106-698-1866	Sequence 1866, Ap
	26	88.6	8.2	517	15	US-10-198-846-12472	Sequence 12472, A
	27	88.6	8.2	1050	14	US-10-200-344-11	Sequence 11, Appl
	28	87.6	8.1	1200	15	US-10-200-910-7	Sequence 7, Appli
	29	83.6	7.7	945	15	US-10-200-910-3	Sequence 3, Appli
	30	83.6	7.7	945	15	US-10-200-910-11	Sequence 11, Appl
	31	80.4	7.4	451	15	US-10-060-036-3265	Sequence 3265, Ap
	32	80	7.4	1125	9	US-09-888-615-2	Sequence 2, Appli
	33	80	7.4	1332	16	US-10-176-306-75	Sequence 75, Appl
	34	80	7.4	1603	16	US-10-176-306-73	Sequence 73, Appl
	35	80	7.4	1826	13	US-10-252-157-453	Sequence 453, App
	36	79.2	7.3	878	15	US-10-198-846-4746	Sequence 4746, Ap
	37	78.4	7.3	741	14	US-10-200-344-5	Sequence 5, Appli
c	38	78.4	7.3	1603	16	US-10-176-306-73	Sequence 73, Appl
	39	76.8	7.1	948	9	US-09-888-615-1	Sequence 1, Appli
	40	76.8	7.1	1187	15	US-10-198-846-13457	Sequence 13457, A
	41	72.8	6.7	1341	9	US-09-925-297-69	Sequence 69, Appl
	42	71.6	6.6	1306	9	US-09-923-779-144	Sequence 144, App
	43	71.6	6.6	1306	9	US-09-962-436-557	Sequence 557, App
	44	70.6	6.5	1254	9	US-09-925-297-58	Sequence 58, Appl
	45	70.6	6.5	1260	13	US-10-345-680-66	Sequence 66, Appl

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Query Match 84.2%; Score 909; DB 10; Length 1728;
Best Local Similarity 89.9%; Pred. No. 7.2e-266;
Matches 1033; Conservative 0; Mismatches 5; Indels 111; Gaps 1;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139
```

```
Qy       121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy       181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy       541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy       592 ------------------------------------------------------------ 591

Db       620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy       592 ------------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db       680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy       610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
            |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db       740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy       670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy       730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy       790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy       850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy       910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
            |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy       970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy      1030 TCGTTTACA 1038
            |||||||||
Db      1160 TCGTTTACA 1168
```

```
RESULT 2
US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1272
;   TYPE: DNA
;   ORGANISM: Papio hamadryas
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1269)
US-10-379-836-1

  Query Match             79.1%;  Score 854.6;  DB 13;  Length 1272;
  Best Local Similarity   86.9%;  Pred. No. 2.3e-249;
  Matches  999;  Conservative    0;  Mismatches   39;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGTCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAGGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              || |||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db        121 GTGCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCGGAC 180

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATTGCATCTGATGTCGACAATGTG 240
              ||||||| |||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        181 CTTATTGAGAAGAAAAAACAAGTCCATTTTTTTGTAAATTCATCTGATGTCGACAATGTG 240

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||| |||||||||| |||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCCTGCTGGCAGATGTGGAAGAT 300

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              |||||||||||||||||||||||||||||||||||||||| |||||||||||| | ||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAACTTATAACTGAGAAGTAT 420

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              |||||||||||||| | ||||||| ||||||| ||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCCTATGAGAAGCACCCACTTTATGTT 480

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATGTGGATTGACTGTGGAATC 540

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              |||||||||||||||||||| ||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCCCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTGAA 600

Qy        592 ------------------------------------------------------------ 591

Db        601 TACTACGGGATAATAGGGGAATATACCAATCTTCTGAGGCATGTGGATTTCTATGTTATG 660
```

```
Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                     ||||||||||||||||||
Db        661 CCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              ||||||||||||||||||||||||| |||||||||||||||||||| || |||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCGTTGCATCGGAACAGACCTGAACAGGAACTTTGCG 780

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTTCTCATGCTCGGAAACCTACTGTGGA 840

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              |||||||||||||||||||||||| |||||| ||||||| ||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGCGAAGGCGGTGGCTAATTTCTTGAGAAGAAATATC 900

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        901 AACCACATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATCGTGTTTCCATAT 960

Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              |||||||| ||||| |||||||||||||| |||||| ||||||||||||||||||||||
Db        961 TCCTATACTCGAAGCAAAAGCAAAGACCACGAGGAATTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||||||| | ||||| |||||||||| |||||||||||||||| ||||||||||
Db       1021 GTTCGTGCTATTCAGAAAACCAGTAAAAATATCAGGTATACACATGGCCGTGGCTCAGAA 1080

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGCGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1141 TCGTTTACA 1149
```

RESULT 3
US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Query Match 40.8%; Score 441; DB 9; Length 1400;
Best Local Similarity 98.9%; Pred. No. 2.7e-123;
Matches 444; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy        590 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 649
              | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        177 AGAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 236

Qy        650 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 709
              |||||||||| |||||||  |||||||||||||||||||||||||||||||||||||||
Db        237 ACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 296

Qy        710 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 769
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        297 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 356

Qy        770 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 829
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        357 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 416

Qy        830 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 889
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        417 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 476

Qy        890 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATA 949
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        477 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATA 536

Qy        950 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 1009
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        537 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 596

Qy       1010 ATGATTTGGGCATCAAATATTCGTTTACA 1038
              |||||||||||||||||||||||||||||
Db        597 ATGATTTGGGCATCAAATATTCGTTTACA 625
```

RESULT 4
US-09-960-352-14595
; Sequence 14595, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14595
; LENGTH: 416
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
US-09-960-352-14595

Query Match 30.7%; Score 331.2; DB 10; Length 416;
Best Local Similarity 87.3%; Pred. No. 3.8e-90;
Matches 363; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy        150 GATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCATTT 209
              |||||||||||||||||| ||| ||||||| |||||||||  | |  ||||||||||
Db          1 GATTGTTCTCTGGCAGCCAGTAGCAGCTGAATATATTGTGAAGGGATACGAAGTCCATTT 60

Qy        210 TTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCC 269
              |||||| ||||||||||||||   |||||||||||||||||||| |||| |||||||
Db         61 TTTTGTGAATGCATCTGATGTAAGCAATGTGAAAGCCCATTTAAATGCGAGCAGAATTCC 120

Qy        270 ATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACAC 329
              || ||| ||| || ||| | | ||||||||||||||| |  ||||||| |||||| |||||
Db        121 ATTCAGGGTCCTGGTGGAAAATGTGGAAGATCTTATCCGGCAGCAGACTTCCAATGACAC 180

Qy        330 AGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTA 389
              |||||||||| || ||| | || |||||||||||||||||||||||||||| |||||
Db        181 CATCAGCCCCCGGGCATCCTCCTCCTACTATGAACAGTATCACTCACTAAATGAGATCTA 240

Qy        390 TTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGG 449
              |||||||||||| |||| |||||| || |||||||||| ||  ||||||||||||||||
Db        241 TTCTTGGATAGAAGTTATGACTGAGCGGTATCCTGATATGGTTGAAAAAATCCACATTGG 300
```

```
Qy        450 ATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGC 509
              ||||||||  ||||||||||||||| ||||||||||||||||||  ||||||||||  ||
Db        301 ATCCTCATACGAGAAGTACCCACTTTATGTTTTAAAGGTTTCTAAAAAAGAACAAAGGGC 360

Qy        510 CAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTG 565
              |||||||||||| ||||||||||||||||||| |||||||| ||||||||||||
Db        361 CAAAAATGCCATGTGGATTGACTGTGGAATCCACGCCAGAGAGTGGATCTCTCCTG 416


RESULT 5
US-09-960-352-8425
; Sequence 8425, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8425
;   LENGTH: 397
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 36-LIB34-005-Q1-E1-A12
US-09-960-352-8425

  Query Match             26.1%;  Score 282.4;  DB 10;  Length 397;
  Best Local Similarity   83.9%;  Pred. No. 2.8e-75;
  Matches  319;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||  ||| | | |||||||||  |||  || || ||||||| | |||||| |
Db         18 ATGAAGCTTTATAGCTTGGGAGTCCTTGTCGCCACCGTCCTGTTCTGTGGGGAGCATGCC 77

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              ||||| || ||||| ||||| ||| || ||||||||||||||||||| ||||||||||||
Db         78 TTCGCCTTTCAGAGGGGCCAGGTTTTATCTGCTCTTCCTAGAACCTCCAGGCAAGTTCAA 137

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||  |||||| |||||||||||||  |||||||||||||||||| ||| |||||||
Db        138 ATTCTGCAGAATGTTACTACAACATACAAGATTGTTCTCTGGCAGCCAGTAGCAGCTGAA 197

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||  | |  |||||||||||||||| ||||||||||||||   |||||||
Db        198 TATATTGTGAAGGGATACGAAGTCCATTTTTTTTGTGAATGCATCTGATGTAAGCAATGTG 257

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||| |||| ||||||||| ||| ||| || ||| | | |||||||||
Db        258 AAAGCCCATTTAAATGCGAGCAGAATTCCATTCAGGGTCCTGGTGGAAAATGTGGAAGAT 317

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||| |  ||||||| |||||| |||||  |||||||||| || ||| | || ||||||
Db        318 CTTATCCGGCAGCAGACTTCCAATGACACCATCAGCCCCCGGGCATCCTCCTCCTACTAT 377

Qy        361 GAACAGTATCACTCACTAAA 380
              ||||||||||||||||||||
Db        378 GAACAGTATCACTCACTAAA 397


RESULT 6
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265133
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265133

  Query Match             12.1%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              |||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 7
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265134
```

```
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134

  Query Match             12.1%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 8
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265135
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265135

  Query Match             12.1%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 9
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265136
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

  Query Match             12.1%;  Score 131;  DB 13;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 10
US-10-027-632-265133
; Sequence 265133, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
```

```
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265133
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265133

  Query Match             12.1%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 11
US-10-027-632-265134
; Sequence 265134, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265134
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265134

  Query Match             12.1%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 12
US-10-027-632-265135
; Sequence 265135, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265135
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265135

  Query Match             12.1%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity   100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547
```

```
RESULT 13
US-10-027-632-265136
; Sequence 265136, Application US/10027632
; GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION:  Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE:  2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  NUMBER OF SEQ ID NOS: 325720
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 265136
;   LENGTH: 1547
;   TYPE: DNA
;   ORGANISM: Human
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1547)
;   OTHER INFORMATION: n = A,T,C or G
US-10-027-632-265136

  Query Match            12.1%;  Score 131;  DB 14;  Length 1547;
  Best Local Similarity  100.0%;  Pred. No. 8.7e-29;
  Matches  131;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        685 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1417 GAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCA 1476

Qy        745 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1477 GAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCA 1536

Qy        805 TACATCAGCAT 815
              ||||||||||||
Db       1537 TACATCAGCAT 1547


RESULT 14
US-09-917-800A-468/c
; Sequence 468, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  APPLICANT:  Gene Logic, Inc.
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5038-US
;  CURRENT APPLICATION NUMBER: US/09/917,800A
;  CURRENT FILING DATE:  2001-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 468
;   LENGTH: 431
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AA998857
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(431)
;   OTHER INFORMATION: n = a or c or g or t
US-09-917-800A-468

  Query Match            12.0%;  Score 129.6;  DB 10;  Length 431;
  Best Local Similarity  87.6%;  Pred. No. 1.1e-28;
  Matches  141;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy        877 CATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAA 936
              |||||||||||||||| |||| ||||| |||||||||||||| ||||| |  | || ||||
Db        431 CATGAGGAACTGTCTNTAGTGGCCAGCGAAGCAGTTCGTGCCATTGAAAGTATTAATAAA 372

Qy        937 AATACCAGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGG 996
              || |||||||| |||||||||  |||||||||||  ||||| ||||||||||||||||
Db        371 AACACCAGGTACACACATGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCT 312

Qy        997 GACGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC 1037
              || ||||||||||||||||||||||||||||||||||||||
Db        311 GATGATTGGATCTATGATTTGGGCATCAAATATTCGTTTAC 271


RESULT 15
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1254
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1254)
US-10-229-546-3

  Query Match            10.0%;  Score 107.6;  DB 15;  Length 1254;
  Best Local Similarity  55.6%;  Pred. No. 1e-21;
  Matches  250;  Conservative    0;  Mismatches  194;  Indels    6;  Gaps    2;
```

```
Qy        590 ATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAG 649
              | || || |||||||||||| || |||||      |     |||    | |||||||| || |
Db        686 AGAACCGCATGTGGAGAAAAAATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTG 745

Qy        650 ACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCAT 709
              |||| || || || |||    |   |   |||           |    ||| |  | |||
Db        746 ACCTCAACAGGAATTTTAATGCTTCA---TGGAACTCCATTCCTAACACCAATGACCCAT 802

Qy        710 GCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTA 769
              |  | || | |||  | ||   |   || ||||| ||   |||   ||| || ||  |||
Db        803 GTGCAGATAACTATCGGGGCTCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTA 862

Qy        770 GTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCC 829
               |||| | |||||  |  | ||  | || || |  ||||||| | | ||||| |||||||
Db        863 ATTTCATTAGAAGCCACCTGAATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCC 922

Qy        830 AGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGT 889
              ||    || ||||||| |||   ||||||  ||            ||||||||||  ||
Db        923 AGATGCTATTGTTTCCCTATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGG 982

Qy        890 CTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATA 949
              |   ||| || |   |     |    || | ||     |||| |  |  | ||| | || |
Db        983 CCAAAGTTGCAA---AGATTGGCACTGATGTTCTATCAACTCGATATGAAACCCGCTACA 1039

Qy        950 CACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCT 1009
                ||||||       |  |||  | ||||      | ||  |     || |||   |
Db       1040 TCTATGGCCCAATAGAATCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTT 1099

Qy       1010 ATGATTTGGGCATCAAATATTCGTTTACAT 1039
              ||||  |||||||||||| |  | ||| | |
Db       1100 ATGACCTGGGCATCAAACACACATTTGCCT 1129
```

Search completed: January 1, 2004, 18:40:43
Job time : 379.077 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 1, 2004, 10:44:24 ; Search time 2497.07 Seconds
(without alignments)
10511.864 Million cell updates/sec

Title: US-09-980-881A-1_COPY_18_1097
Perfect score: 1080
Sequence: 1 atgaagctttgcagccttgc.....ttttgccgctgtctctaaaa 1080

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	660.2	61.1	971	9	AL536069	AL536069 AL536069
2	643.4	59.6	1446	11	AK004045	AK004045 Mus muscu
3	587.2	54.4	788	10	BG618629	BG618629 602646064
4	577.8	53.5	702	2	HSM089979	Bx496754 Homo sapi
5	564.2	52.2	747	10	BG566051	BG566051 602582652
6	562.2	52.1	725	10	BG618239	BG618239 602645967
7	561.8	52.0	738	10	BG616456	BG616456 602614618
8	555.2	51.4	735	9	AV646979	AV646979 AV646979
9	536.2	49.6	541	14	CB157619	CB157619 K-EST0216
10	523.2	48.4	750	10	BG618813	BG618813 602646186
11	508.8	47.1	691	10	BG568240	BG568240 602587149
12	501.2	46.4	601	9	AV658390	AV658390 AV658390
13	494.4	45.8	583	9	AV693037	AV693037 AV693037
14	494.4	45.8	623	9	AV694058	AV694058 AV694058
15	493.4	45.7	602	9	AV692032	AV692032 AV692032
16	492.4	45.6	621	9	AV698425	AV698425 AV698425
17	483.4	44.8	485	13	BX091627	BX091627 BX091627
18	465	43.1	603	10	AW950700	AW950700 EST362770
19	457.8	42.4	537	14	N98450	N98450 za28c09.r1
20	441.6	40.9	623	9	AV658178	AV658178 AV658178
21	441.4	40.9	688	9	AV651709	AV651709 AV651709
22	436	40.4	510	14	CB156851	CB156851 K-EST0215
23	436	40.4	577	14	CB154355	CB154355 K-EST0212
24	435.6	40.3	806	10	BG567264	BG567264 602589745
25	433.4	40.1	462	9	AV681868	AV681868 AV681868
26	427.4	39.6	455	14	W88434	W88434 zh69f04.r1
c 27	425.2	39.4	973	9	AL536068	AL536068 AL536068
28	423.6	39.2	1004	14	BY704293	BY704293 BY704293
29	420.6	38.9	747	10	BG565100	BG565100 602583720
30	417	38.6	853	12	BI144644	BI144644 602909932
31	414.4	38.4	852	10	BG569281	BG569281 602588573
32	409.2	37.9	1022	12	BI332659	BI332659 602984408
33	408.2	37.8	958	9	AI322676	AI322676 mj24d06.y
34	399.6	37.0	434	14	CB164014	CB164014 K-EST0225
35	396	36.7	580	9	AI182113	AI182113 ud73f10.y
36	396	36.7	589	9	AA968307	AA968307 uc71h07.y
37	395	36.6	544	10	BE235838	BE235838 143448 MA
38	394.4	36.5	661	9	AI529872	AI529872 ui83d11.y
39	392.8	36.4	724	9	AI048167	AI048167 ud71b12.y
40	383.4	35.5	615	9	AV658128	AV658128 AV658128
41	382.2	35.4	511	9	AI528146	AI528146 ui94a04.y
42	381.6	35.3	511	9	AI574117	AI574117 uj68d09.y
43	372.8	34.5	610	9	AI574064	AI574064 uj67g11.y
44	369.6	34.2	854	10	BE971100	BE971100 601651336
45	367.8	34.1	787	14	CB593837	CB593837 AGENCOURT

ALIGNMENTS

RESULT 1
AL536069
LOCUS AL536069 971 bp mRNA linear EST 31-MAY-2003
DEFINITION AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION AL536069
VERSION AL536069.2 GI:31260939
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 971)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799562.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6944.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AC11QP1&cluster=6944.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :

```
          http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
          Faraday Avenue Genoscope sequence ID : CS0DF022AC11QP1.
FEATURES             Location/Qualifiers
     source          1. .971
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE21"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      282 a    201 c    211 g    272 t      5 others
ORIGIN

  Query Match             61.1%;  Score 660.2;  DB 9;  Length 971;
  Best Local Similarity   86.7%;  Pred. No. 2.5e-174;
  Matches  783;  Conservative    3;  Mismatches    6;  Indels  111;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 AWGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 124

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 184

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 244

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 304

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 364

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 424

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 484

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 544

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        545 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 604

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT--------- 591
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        605 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 664

Qy        592 ------------------------------------------------------------ 591

Db        665 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 724

Qy        592 ----------------------------------------AATCGAATGTGGAGAAAG 609
                                                       ||||||||||||||||||
Db        725 CCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 784

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||||||||||||||||||||||||||| |||||||||||||||| |||||||

Db        785 AACCGTTCTTTCTATGCGAACAATCATTGCATTGGAACAGACCTGAATAGGAACTTTGCT 844

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        845 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 904

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||:
Db        905 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAWATM 964

Qy        790 AAC 792
              |||
Db        965 AAC 967


RESULT 2
AK004045
LOCUS       AK004045               1446 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 18-day embryo whole body cDNA, RIKEN full-length
            enriched library, clone:1110032P04 product:carboxypeptidase B2
            (plasma), full insert sequence.
ACCESSION   AK004045
VERSION     AK004045.1  GI:12835067
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
```

Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1446)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1446
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1110032P04"
/db_xref="MGI:1900711"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
CDS 42. .1310
/note="unnamed protein product; carboxypeptidase B2 (plasma) (MGD|MGI:1891837, GB|NM_019775, evidence: BLASTN, 99%, match=1419)
putative"
/codon_start=1
/protein_id="BAB23141.1"
/db_xref="GI:12835068"
/db_xref="MGI:1891837"
/translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT 421 a 329 c 311 g 385 t
ORIGIN

Query Match 59.6%; Score 643.4; DB 11; Length 1446;
Best Local Similarity 76.1%; Pred. No. 1.6e-169;
Matches 874; Conservative 0; Mismatches 161; Indels 114; Gaps 2;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||    ||||||  | ||||  ||| |||    |  || ||| ||||||||||| |
Db         42 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCA---TCATCCTCTATGAGCAGCATGGC 98

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||  ||  ||||||||||| |||  ||  |||||||||| ||||||| ||||||||||||
Db         99 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 158

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              | ||  ||||||||||||||||||  ||||||  | ||||||||||||||  ||  ||||||||
Db        159 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 218

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              | ||  | |||||||||  |||||||  ||||||||  |||||  |||||||||||||  |||
Db        219 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 278

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||  ||||||||||||||||  |||||||||  |  ||  ||  ||     |||||||  ||
Db        279 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 338

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              || ||| |||||||||  || |||  |||||  ||||||||||  ||||||||  ||  ||||||
Db        339 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 398

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              || ||||||||||| ||||||||||||||||||  |||||||||  | ||||||||    ||||
Db        399 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 458

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              |||||  |||||     ||||||  |||| |||||  |||||  ||||||||||||||  ||||||
Db        459 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 518

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              ||||||||  || |||||  |||||||  |  ||||||||||||  |||||  ||||||||||||
Db        519 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 578

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGC------------ 588
              ||||||||||||||||  || ||||||||||| ||||||||||||||
Db        579 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 638

Qy        589 ------------------------------------------------------------ 588

Db        639 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 698

Qy        589 ----------------------------------------CATAATCGAATGTGGAGAAAG 609
                                                       | ||||||||||||| |||
Db        699 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 758

Qy        610 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTTGTC 669
              |||||  |||     |   ||||||  |  |||  | ||  |||||||||| || ||||| |
Db        759 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 818

Qy        670 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 729
              |||||||||||||||||  |||||||  ||  ||||||||  |||||  ||||||||||||||
Db        819 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 878

Qy        730 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 789
              ||||||||||||||  || |||||  ||||||||||||||    ||||||||||||||||
Db        879 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 938

Qy        790 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 849
              ||||  |||||||| ||||||||  |||||  |||||||||||  || |||  ||||||| |||
Db        939 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 998
```

```
Qy        850 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 909
              |||||||   |||| ||||||||| ||||| || ||||||||||||| ||||| ||||||
Db        999 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1058

Qy        910 GTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 969
              |||||||| ||||| |  | || |||||| |||||||| ||||| |||  ||||||||||
Db       1059 GTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1118

Qy        970 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1029
              |  ||||| ||||||||||||||   ||||||||||||||||||||||||||||||||||
Db       1119 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1178

Qy       1030 TCGTTTACA 1038
              |||||||||
Db       1179 TCGTTTACA 1187


RESULT 3
BG618629
LOCUS       BG618629                788 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
            mRNA sequence.
ACCESSION   BG618629
VERSION     BG618629.1  GI:13670000
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: o  column: 15
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767974"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      222 a    157 c    169 g    240 t
ORIGIN

  Query Match           54.4%;  Score 587.2;  DB 10;  Length 788;
  Best Local Similarity   99.5%;  Pred. No. 7.8e-154;
  Matches  589;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 125

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 185

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 245

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 305

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        306 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 365

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 425

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 485

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 545

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        546 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 605

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 592
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 657


RESULT 4
HSM089979
ID   HSM089979  standard; RNA; EST; 702 BP.
XX
AC   BX496754;
XX
SV   BX496754.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp779I1529_r1  (from clone DKFZp779I1529)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-702
RA   Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by EMBL (European Molecular Biology Laboratories,
CC   Heidelberg/Germany) within the cDNA sequencing consortium of
CC   the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp779I1529) is available at the RZPD in Berlin.
```

```
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .702
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp779I1529"
FT                   /clone_lib="779 (synonym: hncc1). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="fetal"
FT                   /tissue_type="liver"
XX
SQ   Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;

  Query Match            53.5%;  Score 577.8;  DB 2;  Length 702;
  Best Local Similarity  99.5%;  Pred. No. 3.3e-151;
  Matches  590;  Conservative    0;  Mismatches    2;  Indels    1;  Gaps    1;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         59 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 118

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 178

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 238

Qy        181 CTTATTGTGAAG-AAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGT 239
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        239 CTTATTGTGAAGAAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGT 298

Qy        240 GAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGA 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 GAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGA 358

Qy        300 TCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTA 359
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTA 418

Qy        360 TGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCA 419
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 TGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCA 478

Qy        420 TCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGT 479
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 TCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGT 538

Qy        480 TTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAAT 539
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        539 TTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAAT 598

Qy        540 CCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 651


RESULT 5
BG566051
LOCUS       BG566051                747 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602582652F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
            mRNA sequence.
ACCESSION   BG566051
VERSION     BG566051.1  GI:13573704
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1549  row: g  column: 13
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4710564"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    157 c    159 g    219 t
ORIGIN

  Query Match            52.2%;  Score 564.2;  DB 10;  Length 747;
  Best Local Similarity  99.2%;  Pred. No. 2.3e-147;
  Matches  588;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              |||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db         21 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTG-TCTCTTCTGTGAGCAGCATGTC 79

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTATAACCTCTAGGCAAGTTCAA 139

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499
```

```
Qy        481 T-TAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAAT 539
              | ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        500 TCTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAAT 559

Qy        540 CCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 612


RESULT 6
BG618239
LOCUS       BG618239                725 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
            mRNA sequence.
ACCESSION   BG618239
VERSION     BG618239.1  GI:13669610
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1629  row: f  column: 17
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767376"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      203 a    153 c    149 g    220 t
ORIGIN

  Query Match           52.1%;  Score 562.2;  DB 10;  Length 725;
  Best Local Similarity   99.2%;  Pred. No. 8.2e-147;
  Matches  586;  Conservative    0;  Mismatches    3;  Indels    2;  Gaps    2;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 124

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 184

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db        185 GTTCTACAGAATCTTACTACAACATATGAGA-TGTTCTCTGGCAGCCGGTAACAGCTGAC 243

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 CTTA-TGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 302

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db        303 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 362

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 422

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 482

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 542

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        543 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 602

Qy        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 591
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 653


RESULT 7
BG616456
LOCUS       BG616456                738 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
            mRNA sequence.
ACCESSION   BG616456
VERSION     BG616456.1  GI:13667827
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1595  row: c  column: 05
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4733500"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
```

CA). Note: this is a NIH_MGC Library."
BASE COUNT 209 a 161 c 153 g 215 t
ORIGIN

Query Match 52.0%; Score 561.8; DB 10; Length 738;
Best Local Similarity 99.6%; Pred. No. 1.1e-146;
Matches 563; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy         28 GTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGAGTGGCCAAGTTCTA 87
              ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db          1 GTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTA 60

Qy         88 GCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATAT 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATAT 120

Qy        148 GAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCAT 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCAT 180

Qy        208 TTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATT 267
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATT 240

Qy        268 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGAC 327
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGAC 300

Qy        328 ACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATC 387
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATC 360

Qy        388 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 447
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 420

Qy        448 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACA 507
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db        421 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCA 480

Qy        508 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 567
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 540

Qy        568 TTCTGCTTGTGGTTCATAGGCCATA 592
              |||||||||||||||||||||||||
Db        541 TTCTGCTTGTGGTTCATAGGCCATA 565
```

RESULT 8
AV646979
LOCUS AV646979 735 bp mRNA linear EST 15-JAN-2002
DEFINITION AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.
ACCESSION AV646979
VERSION AV646979.1 GI:9867993
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 735)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES Location/Qualifiers
source 1. .735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GLCATD06"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/clone_lib="GLC"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 200 a 158 c 172 g 205 t
ORIGIN

Query Match 51.4%; Score 555.2; DB 9; Length 735;
Best Local Similarity 98.4%; Pred. No. 7.7e-145;
Matches 571; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

```
Qy         14 GCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTCCAGA 73
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||| ||||
Db          5 GCCTTGCGGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTTCAGA 64

Qy         74 GTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATC 133
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 GTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATC 124

Qy        134 TTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGA 193
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        125 TTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGGTGACCTTATTGTGAAGA 184

Qy        194 AAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAA 253
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 AAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAA 244

Qy        254 ATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGC 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 ATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGC 304

Qy        314 AGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACT 373
              |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        305 AGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGGACTATGAACAGTATCACT 364

Qy        374 CACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTA 433
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 CACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTA 424

Qy        434 CAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAA-GGTTTCT 492
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||
Db        425 CAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGGTGTCT 484

Qy        493 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 552
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAA 544

Qy        553 TGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATA 592
              |||||||||||||||||| ||| || ||||||||||||
Db        545 TGGATCTCTCCTGCTTTCTGGTTGGGGGTCATAGGCCATA 584
```

RESULT 9
CB157619
LOCUS CB157619 541 bp mRNA linear EST 29-JAN-2003
DEFINITION K-EST0216807 L17N670205n1 Homo sapiens cDNA clone L17N670205n1-12-A11 5', mRNA sequence.
ACCESSION CB157619
VERSION CB157619.1 GI:28142751

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 12  row: A  column: 11
            High quality sequence stop: 541.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L17N670205n1-12-A11"
                     /sex="F"
                     /lab_host="Top10F'"
                     /clone_lib="L17N670205n1"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
BASE COUNT      156 a    121 c    107 g    157 t
ORIGIN

  Query Match             49.6%;  Score 536.2;  DB 14;  Length 541;
  Best Local Similarity   99.4%;  Pred. No. 1.5e-139;
  Matches  538;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          5 AGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCG 64
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCG 60

Qy         65 CGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTC 124
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTC 120

Qy        125 TACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTA 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTA 180

Qy        185 TTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAG 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAG 240

Qy        245 CCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTA 304
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        241 CCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTA 300

Qy        305 TTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAAC 364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAAC 360

Qy        365 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTG 424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTG 420

Qy        425 ATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAA 484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAA 480

Qy        485 AGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATG 544
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        481 AGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATG 540

Qy        545 C 545
              |
Db        541 C 541


RESULT 10
BG618813
LOCUS       BG618813                 750 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
            mRNA sequence.
ACCESSION   BG618813
VERSION     BG618813.1  GI:13670184
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: k  column: 22
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767885"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      219 a    146 c    171 g    213 t      1 others
ORIGIN

  Query Match             48.4%;  Score 523.2;  DB 10;  Length 750;
  Best Local Similarity   97.9%;  Pred. No. 7.8e-136;
  Matches  572;  Conservative    0;  Mismatches    8;  Indels    4;  Gaps    4;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 70

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 130

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 190
```

```
Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 250

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 310

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db        311 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGACCCCGAGCCTCCGCATCGTACTAT 370

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 430

Qy        421 CCTGATATGCTT-ACAAAAATCCACATTGG-ATCCTCATTTGAGAAGTACCCACTC-TAT 477
              |||||||||||| |||||||||||||||| |||||||| |||||||| ||| || |||
Db        431 CCTGATATGCTTAACAAAAATCCACATTGGCATCCTCATTCGAGAAGTAACCAGTCGTAT 490

Qy        478 GTTTTAAAGGTTTCT-GGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGG 536
              |||||||||||| || |||||||||||| ||| ||||||||||||||||||||||||||||
Db        491 GTTTTAAAGGTTCCTGGGAAAAGAACAAGCAGGCAAAAATGCCATATGGATTGACTGTGG 550

Qy        537 AATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGT 580
              |||||||||||||||||||||||||||||||||||||||||||||
Db        551 AATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGT 594
```

RESULT 11
BG568240
LOCUS BG568240 691 bp mRNA linear EST 10-APR-2001
DEFINITION 602587149F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4715797 5', mRNA sequence.
ACCESSION BG568240
VERSION BG568240.1 GI:13575893
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 691)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1563 row: a column: 14
High quality sequence stop: 691.
FEATURES Location/Qualifiers
source 1. .691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4715797"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."
BASE COUNT 204 a 143 c 144 g 200 t
ORIGIN

Query Match 47.1%; Score 508.8; DB 10; Length 691;
Best Local Similarity 98.5%; Pred. No. 8.5e-132;
Matches 524; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

```
Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         19 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 78

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         79 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 138

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        139 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 198

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        199 CTTA-TGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497

Qy        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACT 532
              |||||||||||||||||||||||| ||||||||||||||||  | | | |
Db        498 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATAAACTCAAT 549
```

RESULT 12
AV658390
LOCUS AV658390 601 bp mRNA linear EST 16-JAN-2002
DEFINITION AV658390 GLC Homo sapiens cDNA clone GLCFMA03 3', mRNA sequence.
ACCESSION AV658390
VERSION AV658390.1 GI:9879404
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 601)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X., Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W., Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X., Hu,G., Gu,J., Chen,Z. and Han,Z.
TITLE Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
MEDLINE 21625106
PUBMED 11752456
COMMENT Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
Tel: 86-21-50801919(ex.45)

```
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES              Location/Qualifiers
     source           1. .601
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="GLCFMA03"
                      /tissue_type="corresponding non cancerous liver tissue"
                      /dev_stage="Adult"
                      /lab_host="SOLR"
                      /clone_lib="GLC"
                      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                      XhoI"
BASE COUNT      167 a    136 c    118 g    177 t      3 others
ORIGIN

  Query Match             46.4%;  Score 501.2;  DB 9;  Length 601;
  Best Local Similarity   99.0%;  Pred. No. 1.1e-129;
  Matches  503;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          4 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 63
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 60

Qy         64 GCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 123
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 120

Qy        124 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 180

Qy        184 ATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ATTGTGAAGAAAAAACAAGTCCATTTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 240

Qy        244 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 300

Qy        304 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 363
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 360

Qy        364 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCT 423
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        361 CAGTATCACTCACTAAATGAAATCTACTCTTGGATAGAATTTATAACTGAGAGGCATCCT 420

Qy        424 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 483
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db        421 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTNTA 480

Qy        484 AAGGTTTCTGGAAAAGAACAAACAGCCA 511
              |||||||||||| ||||||| |||||||
Db        481 AAGGTTTCTGGANAAGAACATACAGCCA 508


RESULT 13
AV693037
LOCUS       AV693037                 583 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV693037 GKC Homo sapiens cDNA clone GKCAFH12 5', mRNA sequence.
ACCESSION   AV693037
VERSION     AV693037.1  GI:10294900
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 583)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES              Location/Qualifiers
     source           1. .583
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="GKCAFH12"
                      /tissue_type="hepatocellular carcinoma"
                      /dev_stage="Adult"
                      /lab_host="SOLR"
                      /clone_lib="GKC"
                      /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                      XhoI"
BASE COUNT      161 a    130 c    116 g    176 t
ORIGIN

  Query Match             45.8%;  Score 494.4;  DB 9;  Length 583;
  Best Local Similarity   98.8%;  Pred. No. 8.8e-128;
  Matches  498;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         16 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 75

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         76 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 135

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        136 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 195

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        196 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 255

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        256 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 315

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        316 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 375

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        376 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 435

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        436 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 495

Qy        481 TTAAAGGTTTCTGGAAAAGAACAA 504
              |||||||||||||| ||  | | |
Db        496 TTAAAGGTTTCTGGGAACCAGCCA 519


RESULT 14
```

```
AV694058
LOCUS       AV694058                623 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV694058 GKC Homo sapiens cDNA clone GKCADB09 5', mRNA sequence.
ACCESSION   AV694058
VERSION     AV694058.1  GI:10295921
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 623)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .623
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GKCADB09"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GKC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      166 a    137 c    130 g    186 t      4 others
ORIGIN

  Query Match            45.8%;  Score 494.4;  DB 9;  Length 623;
  Best Local Similarity  98.8%;  Pred. No. 9.1e-128;
  Matches  498;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 316

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        481 TTAAAGGTTTCTGGAAAAGAACAA 504
              |||||||||||||| ||  | | |
Db        497 TTAAAGGTTTCTGGGAACCAGCCA 520


RESULT 15
AV692032
LOCUS       AV692032                602 bp    mRNA    linear   EST 16-JAN-2002
DEFINITION  AV692032 GKC Homo sapiens cDNA clone GKCACD01 5', mRNA sequence.
ACCESSION   AV692032
VERSION     AV692032.1  GI:10293895
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 602)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .602
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GKCACD01"
                     /tissue_type="hepatocellular carcinoma"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GKC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      165 a    133 c    120 g    182 t      2 others
ORIGIN

  Query Match            45.7%;  Score 493.4;  DB 9;  Length 602;
  Best Local Similarity  98.6%;  Pred. No. 1.7e-127;
  Matches  497;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         61 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120
              |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196
```

```
Qy        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 316

Qy        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGNT 496

Qy        481 TTAAAGGTTTCTGGAAAAGAACAA 504
              |||||||||||||| ||  | | |
Db        497 TTAAAGGTTTCTGGGAACCAGCCA 520
```

Search completed: January 1, 2004, 15:37:50
Job time : 2504.07 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:21:46 ; Search time 22.1712 Seconds
(without alignments)
687.014 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1817	95.1	386	4	US-09-813-133A-2	Sequence 2, Appli
2	1799.5	94.2	423	1	US-07-649-591B-3	Sequence 3, Appli
3	1799.5	94.2	423	1	US-08-277-540-3	Sequence 3, Appli
4	1799.5	94.2	423	1	US-08-430-787A-3	Sequence 3, Appli
5	1799.5	94.2	423	2	US-08-869-057-2	Sequence 2, Appli
6	1788.5	93.6	423	4	US-09-813-133A-4	Sequence 4, Appli
7	640.5	33.5	404	1	US-08-696-139-2	Sequence 2, Appli
8	639	33.4	415	2	US-08-860-882A-57	Sequence 57, Appl
9	639	33.4	415	4	US-09-011-769A-39	Sequence 39, Appl
10	631.5	33.0	424	4	US-09-011-769A-56	Sequence 56, Appl
11	628.5	32.9	417	1	US-07-649-591B-7	Sequence 7, Appli
12	628.5	32.9	417	1	US-08-277-540-7	Sequence 7, Appli
13	628.5	32.9	417	1	US-08-430-787A-7	Sequence 7, Appli
14	623	32.6	716	3	US-09-171-945-125	Sequence 125, App
15	619.5	32.4	417	1	US-07-649-591B-6	Sequence 6, Appli
16	619.5	32.4	417	1	US-08-277-540-6	Sequence 6, Appli
17	619.5	32.4	417	1	US-08-430-787A-6	Sequence 6, Appli
18	604	31.6	307	2	US-08-782-760-6	Sequence 6, Appli
19	604	31.6	307	5	PCT-US96-00995-6	Sequence 6, Appli
20	604	31.6	437	4	US-09-675-305-10	Sequence 10, Appl
21	603.5	31.6	396	1	US-07-649-591B-4	Sequence 4, Appli
22	603.5	31.6	396	1	US-08-277-540-4	Sequence 4, Appli
23	603.5	31.6	396	1	US-08-430-787A-4	Sequence 4, Appli
24	576	30.1	306	1	US-08-696-139-4	Sequence 4, Appli
25	561	29.4	329	4	US-09-011-769A-51	Sequence 51, Appl
26	561	29.4	349	4	US-09-011-769A-47	Sequence 47, Appl
27	554	29.0	349	4	US-09-011-769A-60	Sequence 60, Appl
28	553	28.9	349	4	US-09-011-769A-64	Sequence 64, Appl
29	553	28.9	417	1	US-07-649-591B-8	Sequence 8, Appli
30	553	28.9	417	1	US-08-277-540-8	Sequence 8, Appli
31	553	28.9	417	1	US-08-430-787A-8	Sequence 8, Appli
32	546	28.6	399	4	US-09-710-099-8	Sequence 8, Appli
33	543	28.4	613	3	US-09-171-945-113	Sequence 113, App
34	528.5	27.7	350	4	US-09-675-305-12	Sequence 12, Appl
35	528.5	27.7	419	1	US-07-649-591B-5	Sequence 5, Appli
36	528.5	27.7	419	1	US-08-277-540-5	Sequence 5, Appli
37	528.5	27.7	419	1	US-08-430-787A-5	Sequence 5, Appli
38	519.5	27.2	436	4	US-09-710-099-6	Sequence 6, Appli
39	513	26.8	417	3	US-08-640-906-4	Sequence 4, Appli
40	513	26.8	417	4	US-09-395-936-4	Sequence 4, Appli
41	512	26.8	417	3	US-08-640-906-18	Sequence 18, Appl
42	512	26.8	417	4	US-09-395-936-18	Sequence 18, Appl
43	487.5	25.5	419	3	US-08-640-906-2	Sequence 2, Appli
44	487.5	25.5	419	4	US-09-395-936-2	Sequence 2, Appli
45	482.5	25.2	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Query Match 95.1%; Score 1817; DB 4; Length 386;
Best Local Similarity 98.6%; Pred. No. 1e-188;
Matches 341; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSET 240
              |||||||||||||||||||||||||||||||||||||||||||||| || ||||||||||||
Db        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSET 240

Qy        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300

Qy        301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||:| |||||||||||||||||||||||||||||||||||||
Db        301 SEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
```

RESULT 2

```
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-3

  Query Match             94.2%;  Score 1799.5;  DB 1;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 9.7e-187;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              |||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383
```

```
RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-3

  Query Match             94.2%;  Score 1799.5;  DB 1;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 9.7e-187;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300
```

```
Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 4
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-3

  Query Match             94.2%;  Score 1799.5;  DB 1;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 9.7e-187;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH----------------------------------NRMWRK 203
              |||||||||||||||||                                  ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              ||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  Plasma
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  23..401
US-08-869-057-2

  Query Match             94.2%;  Score 1799.5;  DB 2;  Length 423;
  Best Local Similarity   89.6%;  Pred. No. 9.7e-187;
  Matches  343;  Conservative    1;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
```

```
Qy          61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy         121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy         181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
               |||||||||||||||||                                   ||||||
Db         181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy         204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
               ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db         241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy         264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
               ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db         301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy         324 TLYLAPGGGDDWIYDLGIKYSFT 346
               |||||||||||||||||||||||
Db         361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 6
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001173
;  CURRENT APPLICATION NUMBER: US/09/813,133A
;  CURRENT FILING DATE:  2001-06-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Human
US-09-813-133A-4

  Query Match             93.6%;  Score 1788.5;  DB 4;  Length 423;
  Best Local Similarity   89.0%;  Pred. No. 1.5e-185;
  Matches  341;  Conservative    1;  Mismatches    4;  Indels   37;  Gaps    1;

Qy           1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy          61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy         121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
               |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db         121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGI 180

Qy         181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
               |||||||||||||||||                                   ||||||
Db         181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy         204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
               ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db         241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy         264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
               ||||||||||||||||||||||||||||||||||||||||||:| ||||||||||||||||
Db         301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSE 360

Qy         324 TLYLAPGGGDDWIYDLGIKYSFT 346
               |||||||||||||||||||||||
Db         361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 7
US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  404 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-696-139-2

  Query Match             33.5%;  Score 640.5;  DB 1;  Length 404;
  Best Local Similarity   37.7%;  Pred. No. 7.6e-61;
  Matches  136;  Conservative   60;  Mismatches  122;  Indels   43;  Gaps    6;

Qy          23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
               |:  :|           : :|  | :| :|  |:| :   |     | | | | |:  |:
Db          10 FEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 69

Qy          83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
                 |  : :   ||: ::  ::: |  :     | :   ||:|::   | :| : :|  :||
Db          70 FLEQNELQYEVLINNLRSVLEAQFDSRV---RTTGHSYEKYNNWETIEAWTKQVTSENPD 126

Qy         143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
               ::::  ||::|    :|:||| ||     | ||::||| ||||||| ||| ||:
Db         127 LISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCGFHAREWISHAFCQWFVREAVLTY 185

Qy         196 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                    |:      |||||| || |   ||||| | ||
Db         186 GYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGTTCIGTDPNRNF-DA 244
```

```
Qy        226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
              ||  |||:  | |||||   ||| | ||:| |:| |:: ||||:::|||||  |::|||| 
Db        245 GWCTTGASTDPCDETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 304

Qy        286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
                  ::: ||: :|  ||: :  |    |:||:| |: |:| | || ||| || ||||||
Db        305 DYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGATTIYPAAGGSDDWAYDQGIKYSF 363

Qy        346 T 346
              |
Db        364 T 364


RESULT 8
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  415 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-860-882A-57

  Query Match             33.4%;  Score 639;  DB 2;  Length 415;
  Best Local Similarity   35.6%;  Pred. No. 1.2e-60;
  Matches  135;  Conservative   69;  Mismatches  131;  Indels   44;  Gaps    7;

Qy          6 LAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK 64
              | ||| : |     |    |:  :|          : ::: | :| :|  |:| :   |
Db          2 LLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKP 61

Qy         65 KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYH 124
                  | | | | |   |:  |  : :   ||:::: :::: |  :     ||:   ||:|:

Db         62 HSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYN 118

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
                  | :| : :   :| ::::  ||::||   :|:||| ||     | ||::||| ||||
Db        119 KWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHARE 177

Qy        185 WISPAFCLWFI------------------------------GH------NRMWRKNRSF 207
              ||||||| ||:                              |:      :| ||| ||
Db        178 WISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRST 237

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              :  : ||||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  :: ||
Db        238 HTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIK 296

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              ||:::|||| :::||||     ::: ||: :|   |: :  :   |:||:| |: |:|
Db        297 AYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYP 355

Qy        328 APGGGDDWIYDLGIKYSFT 346
              | || ||| || ||:||||
Db        356 AAGGSDDWAYDQGIRYSFT 374


RESULT 9
US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 39:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 415 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

  Query Match             33.4%;  Score 639;  DB 4;  Length 415;
```

```
  Best Local Similarity   35.6%;  Pred. No. 1.2e-60;
  Matches  135;  Conservative   69;  Mismatches  131;  Indels   44;  Gaps    7;

Qy          6 LAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK 64
              | ||| : |    |    |:  :|          : ::: | :| :|  |:| :   |
Db          2 LLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKP 61

Qy         65 KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYH 124
                | | | | |   |:  |  : :   ||:::: :::: |  :     ||:   ||:|:
Db         62 HSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYN 118

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
                | :| : :   :| ::::  ||::||   :|:||| ||    | ||::||| ||||
Db        119 KWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHARE 177

Qy        185 WISPAFCLWFI------------------------------GH------NRMWRKNRSF 207
              ||||||| ||:                             |:      :| ||| ||
Db        178 WISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRST 237

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              :  : ||||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  :: ||
Db        238 HTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIK 296

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              ||:::||||| :::||||     ::: ||: :|   |: :  :   |:||:| |: |:|
Db        297 AYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYP 355

Qy        328 APGGGDDWIYDLGIKYSFT 346
              | || ||| || ||:||||
Db        356 AAGGSDDWAYDQGIRYSFT 374


RESULT 10
US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 56:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 424 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56

  Query Match             33.0%;  Score 631.5;  DB 4;  Length 424;
  Best Local Similarity   35.7%;  Pred. No. 7.9e-60;
  Matches  129;  Conservative   68;  Mismatches  121;  Indels   43;  Gaps    6;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
              |:  :|          : ::: | :| :|  |:| :   |      | | | | |   |:
Db         29 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 88

Qy         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
               |  : :   ||:::: :::: |  :     ||:   ||:|:     | :| : :   :|
Db         89 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 145

Qy        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        146 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 204

Qy        196 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                     |:      :| ||| || :  : |||||  | ||
Db        205 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 263

Qy        226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||||  :::||||
Db        264 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 323

Qy        286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
                    ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        324 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 382

Qy        346 T 346
              |
Db        383 T 383


RESULT 11
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
```

```
;       TELEFAX:  415/952-9881
;       TELEX:  910/371-7168
;   INFORMATION FOR SEQ ID NO:  7:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  417 amino acids
;       TYPE:  AMINO ACID
;       TOPOLOGY:  linear
US-07-649-591B-7

  Query Match             32.9%;  Score 628.5;  DB 1;  Length 417;
  Best Local Similarity   37.0%;  Pred. No. 1.6e-59;
  Matches  140;  Conservative   60;  Mismatches  131;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKK 65
              :||:   :      |   |   :|     :  :   ||:||| : |:  | |     |
Db          7 MAVIYTTLAIAPVH---FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVN 63

Qy         66 KQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYH 124
               | | |:  :   ::: |     |   :|: |::: |::|     | :: | |   | :|:
Db         64 MTVDFRVSEKESQTIQSTLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
              ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::|||||||
Db        121 DWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy        185 WISPAFCLWFI-------GHN----------------------------RMWRKNRSF 207
              |||||||  ||:       | |                            |||||||
Db        180 WISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              |: |||||||  ||     |      ::  |   | |  |||| | ||| :|:|  ::| ||
Db        240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIK 298

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              |||: |||||  ::  || ||      :|::|  ||  |  |:  |   ||| :|  : |:|
Db        299 AYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYK 357

Qy        328 APGGGDDWIYDLGIKYSF 345
              |   ||:||||||::|
Db        358 TSGSSLDWVYDLGIKHTF 375


RESULT 12
US-08-277-540-7
; Sequence 7, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-7

  Query Match             32.9%;  Score 628.5;  DB 1;  Length 417;
  Best Local Similarity   37.0%;  Pred. No. 1.6e-59;
  Matches  140;  Conservative   60;  Mismatches  131;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKK 65
              :||:   :      |   |   :|     :  :   ||:||| : |:  | |     |
Db          7 MAVIYTTLAIAPVH---FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVN 63

Qy         66 KQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYH 124
               | | |:  :   ::: |     |   :|: |::: |::|     | :: | |   | :|:
Db         64 MTVDFRVSEKESQTIQSTLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
              ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::|||||||
Db        121 DWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy        185 WISPAFCLWFI-------GHN----------------------------RMWRKNRSF 207
              |||||||  ||:       | |                            |||||||
Db        180 WISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              |: |||||||  ||     |      ::  |   | |  |||| | ||| :|:|  ::| ||
Db        240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIK 298

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              |||: |||||  ::  || ||      :|::|  ||  |  |:  |   ||| :|  : |:|
Db        299 AYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYK 357

Qy        328 APGGGDDWIYDLGIKYSF 345
              |   ||:||||||::|
Db        358 TSGSSLDWVYDLGIKHTF 375


RESULT 13
US-08-430-787A-7
; Sequence 7, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
```

```
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-7

  Query Match             32.9%;  Score 628.5;  DB 1;  Length 417;
  Best Local Similarity   37.0%;  Pred. No. 1.6e-59;
  Matches  140;  Conservative   60;  Mismatches  131;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKK 65
              :||:  :     |   |   :|    :  :   ||:||| : |:  | |      |
Db          7 MAVIYTTLAIAPVH---FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVN 63

Qy         66 KQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYH 124
                | | |:  :   ::: |    |   :|: |::: |::|    | :: | |   | :|:
Db         64 MTVDFRVSEKESQTIQSTLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
              ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::||||||||
Db        121 DWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy        185 WISPAFCLWFI-------GHN-----------------------------RMWRKNRSF 207
              ||||||| ||:       | |                             ||||||||
Db        180 WISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
                |: ||||||| ||    |     ::  |   | |  |||| | ||| :|:| ::| ||
Db        240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIK 298

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              |||: |||| :: || ||     :|::|  ||  |  |:  |   ||| :|  : |:|
Db        299 AYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYK 357

Qy        328 APGGGDDWIYDLGIKYSF 345
                |   ||:||||||::|
Db        358 TSGSSLDWVYDLGIKHTF 375


RESULT 14
US-09-171-945-125
; Sequence 125, Application US/09171945
; Patent No. 6277599
; GENERAL INFORMATION:
;  APPLICANT:  Emery, Stephen
;  APPLICANT:  Copley, Clive Graham
;  APPLICANT:  Edge, Michael Derek
;  TITLE OF INVENTION:  Monoclonal Antibody to CEA, Conjugates Comprising Said
;  TITLE OF INVENTION:  Antibody, and Their Therapeutic Use in an Adept System
;  FILE REFERENCE:  Monoclonal Antibody to CEA
;  CURRENT APPLICATION NUMBER: US/09/171,945
;  CURRENT FILING DATE:  1998-10-29
;  PRIOR APPLICATION NUMBER:  GB9703103.3
;  PRIOR FILING DATE:  1997-02-14
;  PRIOR APPLICATION NUMBER:  GB9609405.7
;  PRIOR FILING DATE:  1996-05-04
;  PRIOR APPLICATION NUMBER:  PCT/GB97/01165
;  PRIOR FILING DATE:  1997-04-29
;  NUMBER OF SEQ ID NOS:  131
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 125
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-125

  Query Match             32.6%;  Score 623;  DB 3;  Length 716;
  Best Local Similarity   34.6%;  Pred. No. 1.5e-58;
  Matches  132;  Conservative   72;  Mismatches  134;  Indels   44;  Gaps    7;

Qy          3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADL 61
              : :| ||| : |     |    |:  :|          : ::: | :| :|  |:| :
Db          1 MLALLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQ 60

Qy         62 IVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYE 121
              |      | | | | |    |:  |  : :   ||:::: :::: |  :     ||:   ||
Db         61 IKPHSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYE 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||::||| |
Db        118 KYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              ||||||||||| ||:                              |:      :| |||
Db        177 AREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKT 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              || :  : ||||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  ::
Db        237 RSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLS 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              ||||:::||||| :::|||    ::: ||: :|   |: :  :   |:|||:| |: |
Db        296 SIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATT 354

Qy        325 LYLAPGGGDDWIYDLGIKYSFT 346
              :| : |   || || ||:||||
Db        355 IYPSAGTSKDWAYDQGIRYSFT 376


RESULT 15
US-07-649-591B-6
; Sequence 6, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-6

  Query Match            32.4%;  Score 619.5;  DB 1;  Length 417;
  Best Local Similarity  34.1%;  Pred. No. 1.5e-58;
  Matches  130;  Conservative   76;  Mismatches  128;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFA---FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLI 62
              : :::|: |       |   |   :|    |: :|  ::::|  | |:  | |     :
Db          1 MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHV 60

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYE 121
                | | |:  :   ::: |: : :   :|: |::: |::|    : :  | |   |
Db         61 AANMMVDFRVSEKESQAIQSALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YA 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||||
Db        118 KYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              ||||:|||||  ||:                             |:      |||||||
Db        177 AREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKN 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||   |: |||||||  || :  |     ::  |:: | |  |||| | ||| :|:| ::|
Db        237 RSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              :|| ||: ||||| ::||| ||     :||:|: ||      :  |   ||| :|    |
Db        296 EIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIEST 354

Qy        325 LYLAPGGGDDWIYDLGIKYSF 345
              :|   |   || ||||||::|
Db        355 IYPISGSSLDWAYDLGIKHTF 375


Search completed: January  1, 2004, 19:33:09
Job time : 24.1712 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:32:17 ; Search time 51.4823 Seconds
(without alignments)
1403.024 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1799.5	94.2	423	12	US-10-379-836-17	Sequence 17, Appl
2	1716.5	89.8	423	12	US-10-379-836-2	Sequence 2, Appli
3	1528	80.0	422	12	US-10-379-836-18	Sequence 18, Appl
4	1499	78.4	422	12	US-10-379-836-16	Sequence 16, Appl
5	803.5	42.0	211	9	US-09-925-302-467	Sequence 467, App
6	631.5	33.0	402	12	US-10-379-836-20	Sequence 20, Appl
7	623	32.6	716	10	US-09-910-059-125	Sequence 125, App
8	619.5	32.4	417	12	US-10-341-434-188	Sequence 188, App
9	619.5	32.4	417	15	US-10-229-546-2	Sequence 2, Appli
10	619.5	32.4	417	15	US-10-229-546-9	Sequence 9, Appli
11	604	31.6	437	12	US-10-274-639-12	Sequence 12, Appl
12	604	31.6	437	14	US-10-200-344-10	Sequence 10, Appl
13	563	29.5	374	9	US-09-888-615-61	Sequence 61, Appl
14	563	29.5	444	16	US-10-176-306-74	Sequence 74, Appl
15	546	28.6	399	15	US-10-200-910-8	Sequence 8, Appli
16	543	28.4	613	10	US-09-910-059-113	Sequence 113, App
17	528.5	27.7	350	14	US-10-200-344-12	Sequence 12, Appl
18	519.5	27.2	436	15	US-10-200-910-6	Sequence 6, Appli
19	511.5	26.8	436	14	US-10-076-535-2	Sequence 2, Appli
20	508	26.6	417	9	US-09-923-779-150	Sequence 150, App
21	504	26.4	428	9	US-09-925-297-528	Sequence 528, App
22	495.5	25.9	298	12	US-10-379-836-19	Sequence 19, Appl
23	489.5	25.6	406	9	US-09-925-297-517	Sequence 517, App
24	489.5	25.6	419	12	US-10-345-680-65	Sequence 65, Appl
25	483	25.3	421	11	US-09-946-374-234	Sequence 234, App
26	483	25.3	421	12	US-10-015-387A-234	Sequence 234, App
27	483	25.3	421	12	US-10-006-130A-234	Sequence 234, App
28	483	25.3	421	12	US-10-199-672-308	Sequence 308, App
29	483	25.3	421	12	US-10-006-172A-234	Sequence 234, App
30	483	25.3	421	12	US-10-187-749-308	Sequence 308, App
31	483	25.3	421	12	US-10-194-457-308	Sequence 308, App
32	483	25.3	421	12	US-10-184-642-308	Sequence 308, App
33	483	25.3	421	12	US-10-196-747-308	Sequence 308, App
34	483	25.3	421	12	US-10-015-392A-234	Sequence 234, App
35	483	25.3	421	12	US-10-017-253A-234	Sequence 234, App
36	483	25.3	421	12	US-10-173-689-308	Sequence 308, App
37	483	25.3	421	12	US-10-173-690-308	Sequence 308, App
38	483	25.3	421	12	US-10-173-691-308	Sequence 308, App
39	483	25.3	421	12	US-10-173-692-308	Sequence 308, App
40	483	25.3	421	12	US-10-173-694-308	Sequence 308, App
41	483	25.3	421	12	US-10-173-698-308	Sequence 308, App
42	483	25.3	421	12	US-10-173-699-308	Sequence 308, App
43	483	25.3	421	12	US-10-173-707-308	Sequence 308, App
44	483	25.3	421	12	US-10-174-569-308	Sequence 308, App
45	483	25.3	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Query Match 94.2%; Score 1799.5; DB 12; Length 423;
Best Local Similarity 89.6%; Pred. No. 2.9e-175;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

```
Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH----------------------------------NRMWRK 203
```

```
                ||||||||||||||||||                                    ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
                ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
                |||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
                |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Papio hamadryas
US-10-379-836-2

  Query Match             89.8%;  Score 1716.5;  DB 12;  Length 423;
  Best Local Similarity   85.4%;  Pred. No. 9e-167;
  Matches  327;  Conservative    7;  Mismatches   12;  Indels   37;  Gaps    1;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              || ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIEKKKQVHFFVNSSDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||| |||::||||||||||||:||:||||||||||||||||:||||||
Db        121 EQYHSLNEIYSWIELITEKYPDMLTKIHIGSSYEKHPLYVLKVSGKEQTAKNAMWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||                                   ||||||
Db        181 HAREWISPAFCLWFIGHITEYYGIIGEYTNLLRHVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||| ||||||| || |||||||||| |||||||||||||||| ||||:||||||
Db        241 NRSFYANNRCIGTDLNRNFASKHWCEEGASSFSCSETYCGLYPESEPEAKAVANFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              | |||||||||||||||||||||||||||||||||||||||||| ||||| ||||| |||
Db        301 NHIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIQKTSKNIRYTHGRGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||| ||||||||||||||
Db        361 TLYLAPGGADDWIYDLGIKYSFT 383


RESULT 3
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-379-836-18

  Query Match             80.0%;  Score 1528;  DB 12;  Length 422;
  Best Local Similarity   76.2%;  Pred. No. 1.8e-147;
  Matches  292;  Conservative   24;  Mismatches   29;  Indels   38;  Gaps    2;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              |||  | :|| |:|: ||| |||||||||:||||||||:|||||||||:||||||||:
Db          1 MKLHGLGILVAIILY-EQHGFAFQSGQVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAE 59

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
                | |||:||||||||||:|||||||| || :||: :|||||:||  ||||||||||||
Db         60 FIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIEQQTFNDTVSPRASASYY 119

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||| |||:||||| ||:|||||||||||||||||  ||||||||||
Db        120 EQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAIWIDCGI 179

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||||:                                   ||||||
Db        180 HAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRK 239

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||| : || |:||||| || ||||||:||||||||||||||||||||||||| ||||||
Db        240 NRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSETYCGLYPESEPEVKAVADFLRRNI 299

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              : ||||||||||| |:||||| |||||||||||||||||||:  :|||||||| |||
Db        300 DHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSE 359

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              :||||||| ||||||||||||||
Db        360 SLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 4
US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-379-836-16

  Query Match             78.4%;  Score 1499;  DB 12;  Length 422;
```

```
  Best Local Similarity   74.7%;  Pred. No. 1.6e-144;
  Matches  286;  Conservative   26;  Mismatches   33;  Indels   38;  Gaps    2;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              |||  | |||  |:|: |:|  ||||| ||:||||||||||:|||||||||:||||||||:
Db          1 MKLYGLGVLVAIILY-EKHGLAFQSGHVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAE 59

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
               | |||:||||||||||::|||:|| | || :||: :|||||||| ||||||||||:|||
Db         60 FIEKKKEVHFFVNASDVNSVKAYLNASRIPFNVLMNNVEDLIQQQTSNDTVSPRASSSYY 119

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||| |||:||||| ||:||||:||||||||||||   ||||||||||
Db        120 EQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSYEKYPLYVLKVSGKEHRVKNAIWIDCGI 179

Qy        181 HAREWISPAFCLWFIGH-------------------------------------NRMWRK 203
              |||||||||||||||:                                     ||||||
Db        180 HAREWISPAFCLWFIGYVTQFHGKENTYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRK 239

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||| : || |:||||| || ||||||:|||| ||||||||||||||||||||| ||||||
Db        240 NRSVHMNNRCVGTDLNRNFASKHWCEKGASSFSCSETYCGLYPESEPEVKAVADFLRRNI 299

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              | |||||||||||| |:||||| |||||||||||||||||||:  :||||||| |||
Db        300 NHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSE 359

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              :|||||| ||||||||||||||
Db        360 SLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 5
US-09-925-302-467
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
;   LENGTH: 211
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-302-467

  Query Match            42.0%;  Score 803.5;  DB 9;  Length 211;
  Best Local Similarity   88.8%;  Pred. No. 9e-74;
  Matches  151;  Conservative    2;  Mismatches    8;  Indels    9;  Gaps    2;

Qy        185 WISPAFCLWFI--------GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSS 236
              |||   | | :          |||||||||||||||||||| || ||||||||||||||
Db          3 WIS-MLCRWLMWMVMNYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 61

Qy        237 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 296
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 121

Qy        297 SLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              ||||||||||:||||||||||||||||||||||||||||||||||||||
Db        122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 171


RESULT 6
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-379-836-20

  Query Match            33.0%;  Score 631.5;  DB 12;  Length 402;
  Best Local Similarity   35.7%;  Pred. No. 9.5e-56;
  Matches  129;  Conservative   68;  Mismatches  121;  Indels   43;  Gaps    6;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |:
Db          7 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 66

Qy         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
              |  : :   ||:::: :::: |  :     ||:   ||:|:     | :| : :   :|
Db         67 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 123

Qy        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        124 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 182

Qy        196 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                     |:      :| ||| || :  : ||||| | ||
Db        183 GREIQVTELLNKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 241

Qy        226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::||||| :::||||
Db        242 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 301

Qy        286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
                  ::: ||: :|    |: :  :    |:||:| |: |:| | || ||| || ||:|||
Db        302 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 360

Qy        346 T 346
              |
Db        361 T 361


RESULT 7
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
```

```
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-125

  Query Match            32.6%;  Score 623;  DB 10;  Length 716;
  Best Local Similarity  34.6%;  Pred. No. 1.7e-54;
  Matches  132;  Conservative   72;  Mismatches  134;  Indels   44;  Gaps    7;

Qy          3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADL 61
              : :| ||| : |     |     |: :|           : ::: | :| :|  |:| :
Db          1 MLALLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQ 60

Qy         62 IVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYE 121
              |     | | | | |    |:  |  : :   ||:::: :::: |  :      ||:   ||
Db         61 IKPHSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYE 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||::||| |
Db        118 KYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              |||||||||| ||:                              |:      :| |||
Db        177 AREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKT 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              || :  : ||||| | ||     ||| ||| : | |||||    ||| | ||:| |:|  ::
Db        237 RSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLS 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              ||||:::||||| :::||||     ::: ||: :|   |: :  :   |:||:| |: |
Db        296 SIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATT 354

Qy        325 LYLAPGGGDDWIYDLGIKYSFT 346
              :| : |   || || ||:||||
Db        355 IYPSAGTSKDWAYDQGIRYSFT 376


RESULT 8
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-341-434-188

  Query Match            32.4%;  Score 619.5;  DB 12;  Length 417;
  Best Local Similarity  34.1%;  Pred. No. 1.7e-54;
  Matches  130;  Conservative   76;  Mismatches  128;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFA---FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLI 62
              : :::|: |       |   |   :|    |:  :|  ::::|  | |:  | |     :
Db          1 MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHV 60

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYE 121
                | | |:  :   ::: |: : :   :|: |::: |::|    : :  | |   |
Db         61 AANMMVDFRVSEKESQAIQSALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YA 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||||
Db        118 KYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              ||||:|||||  ||:                              |:      |||||||
Db        177 AREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKN 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||   |: ||||||| || :  |      ::  |:: | |  |||| | ||| :|:| ::|
Db        237 RSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              :|| ||: ||||| ::|||| ||     :||:|: ||       :  |  ||| :|    |
Db        296 EIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIEST 354

Qy        325 LYLAPGGGDDWIYDLGIKYSF 345
              :|   |   || ||||||::|
Db        355 IYPISGSSLDWAYDLGIKHTF 375


RESULT 9
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-2

  Query Match            32.4%;  Score 619.5;  DB 15;  Length 417;
  Best Local Similarity  34.1%;  Pred. No. 1.7e-54;
  Matches  130;  Conservative   76;  Mismatches  128;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFA---FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLI 62
              : :::|: |       |   |   :|    |:  :|  ::::|  | |:  | |     :
Db          1 MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHV 60

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYE 121
                | | |:  :   ::: |: : :   :|: |::: |::|    : :  | |   |
Db         61 AANMMVDFRVSEKESQAIQSALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YA 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||||
Db        118 KYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              ||||:|||||  ||:                              |:      |||||||
Db        177 AREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKN 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||   |: ||||||| || :  |      ::  |:: | |  |||| | ||| :|:| ::|
Db        237 RSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
```

```
              :|| ||: ||||| ::||| ||     :||:|: ||      :  |   ||| :|    |
Db        296 EIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIEST 354

Qy        325 LYLAPGGGDDWIYDLGIKYSF 345
              :|   |   || ||||||::|
Db        355 IYPISGSSLDWAYDLGIKHTF 375


RESULT 10
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-9

  Query Match             32.4%;  Score 619.5;  DB 15;  Length 417;
  Best Local Similarity   34.1%;  Pred. No. 1.7e-54;
  Matches  130;  Conservative   76;  Mismatches  128;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFA---FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLI 62
              : :::|: |       |   |   :|    |:  :|  ::::|  | |:  | |     :
Db          1 MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHV 60

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYE 121
                | | |:  :   ::: |: : :   :|: |::: |::|    : :  | |   |
Db         61 AANMMVDFRVSEKESQAIQSALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YA 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||||
Db        118 KYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              ||||:|||| ||:                              |:      ||||||
Db        177 AREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKN 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||   |: ||||||| || :  |     ::  |:: | |  |||| | ||| :|:| ::|
Db        237 RSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              :|| ||: ||||| ::||| ||     :||:|: ||      :  |   ||| :|    |
Db        296 EIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIEST 354

Qy        325 LYLAPGGGDDWIYDLGIKYSF 345
              :|   |   || ||||||::|
Db        355 IYPISGSSLDWAYDLGIKHTF 375


RESULT 11
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 12
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

  Query Match             31.6%;  Score 604;  DB 12;  Length 437;
  Best Local Similarity   34.4%;  Pred. No. 7.1e-53;
  Matches  134;  Conservative   67;  Mismatches  135;  Indels   54;  Gaps   10;

Qy          1 MKLC--SLAVLVPIVLFCEQHVF--AFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQP 56
              : ||   | :| |       |::   :   :|:  :|:|  :   |: ::   :: ||||
Db         15 LPLCWLFLKILQP----GHSHLYNNRYAGDKVIRFIPKTEEEAYALKKISYQLKVDLWQP 70

Qy         57 VTADLIVK--KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPR 114
                :   : :     ||   | |    : | |  : |   ||: |::  :::  | |   |
Db         71 SSISYVSEGTVTDVHIPQNGSRA--LLAFLQEANIQYKVLIEDLQKTLEKGSSLHTQRNR 128

Qy        115 ASAS--YYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKN 172
                | |   || |||| || :|:  : : |  ::    || |:|   |::||: |:    |
Db        129 RSLSGYNYEVYHSLEEIQNWMHHLNKTHSGLIHMFSIGRSYEGRSLFILKL-GRRSRLKR 187

Qy        173 AIWIDCGIHAREWISPAFCLWFI---------------------------------- 195
              |:||||||||||||| |||| ||:
Db        188 AVWIDCGIHAREWIGPAFCQWFVKEALLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSW 247

Qy        196 GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAV 255
              ::| ||| ||  :   | | | | |:  | ||:||||   | :|||| :||||||||||
Db        248 TNDRFWRKTRSRNSRFRCRGVDANRNWKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAV 306

Qy        256 ASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTR 315
              |:|||::   |:||:| |:|:| :::|||| :   :   :   | :|| |: ::    |
Db        307 ANFLRKHRKHIRAYLSFHAYAQMLLYPYSYKYATIPNFRCVESAAYKAVNAL-QSVYGVR 365

Qy        316 YTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
              | :|  | |||:: |   || |  || |:|
```

```
Db        366 YRYGPASTTLYVSSGSSMDWAYKNGIPYAF 395


RESULT 12
US-10-200-344-10
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-200-344-10

  Query Match             31.6%;  Score 604;  DB 14;  Length 437;
  Best Local Similarity   34.4%;  Pred. No. 7.1e-53;
  Matches  134;  Conservative   67;  Mismatches  135;  Indels   54;  Gaps   10;

Qy          1 MKLC--SLAVLVPIVLFCEQHVF--AFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQP 56
              : ||   | :| |       |::   :   :|:  :|:|  :   |: ::   :: ||||
Db         15 LPLCWLFLKILQP----GHSHLYNNRYAGDKVIRFIPKTEEEAYALKKISYQLKVDLWQP 70

Qy         57 VTADLIVK--KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPR 114
               :   : :     ||   | |    : | |  : |   ||: |::  :::  |  |   |
Db         71 SSISYVSEGTVTDVHIPQNGSRA--LLAFLQEANIQYKVLIEDLQKTLEKGSSLHTQRNR 128

Qy        115 ASAS--YYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKN 172
              | |   || |||| || :|:  : : |  ::    || |:|   |::||: |:    |
Db        129 RSLSGYNYEVYHSLEEIQNWMHHLNKTHSGLIHMFSIGRSYEGRSLFILKL-GRRSRLKR 187

Qy        173 AIWIDCGIHAREWISPAFCLWFI---------------------------------- 195
              |:||||||||||| |||| ||:
Db        188 AVWIDCGIHAREWIGPAFCQWFVKEALLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSW 247

Qy        196 GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAV 255
              ::| ||| ||  :   | | | | |:  | ||:||||   | :|||| :||||||||||
Db        248 TNDRFWRKTRSRNSRFRCRGVDANRNWKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAV 306

Qy        256 ASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTR 315
              |:||::   |:||:| |:|:| :::|||| :   :   :   | :|| |: ::    |
Db        307 ANFLRKHRKHIRAYLSFHAYAQMLLYPYSYKYATIPNFRCVESAAYKAVNAL-QSVYGVR 365

Qy        316 YTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
              | :|  | |||:: |   || |  || |:|
Db        366 YRYGPASTTLYVSSGSSMDWAYKNGIPYAF 395


RESULT 13
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-888-615-61

  Query Match             29.5%;  Score 563;  DB 9;  Length 374;
  Best Local Similarity   38.7%;  Pred. No. 8.8e-49;
  Matches  109;  Conservative   51;  Mismatches   82;  Indels   40;  Gaps    5;

Qy        103 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVL 161
              :|:| : :||| :  :| |  || : ||| |:  |:|:: :::|:  :| ::| :|:| |
Db         29 RQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL 88

Qy        162 KVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI---------------------------- 195
              |:|      |  ||:||||||||||:|||| ||:
Db         89 KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLP 148

Qy        196 -----GH------NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGL 244
                  |:       :|:|||:|| : |  | ||||| || :  ||  ||| :   :|:||
Db        149 VLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGT 207

Qy        245 YPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAV 304
               | |||| ||||||:    : |  :::|||| | |: || ||::|| :| |:  |  :|
Db        208 GPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAA 267

Qy        305 RAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |: |   | |  |  :: || : |   ||  |:|| :|:|
Db        268 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 308


RESULT 14
US-10-176-306-74
; Sequence 74, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
```

```
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 444
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-306-74

  Query Match             29.5%;  Score 563;  DB 16;  Length 444;
  Best Local Similarity   38.7%;  Pred. No. 1.1e-48;
  Matches  109;  Conservative   51;  Mismatches   82;  Indels   40;  Gaps    5;

Qy        103 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVL 161
              :|:| : :||| :  :| |  || : ||| |:  |:|:: :::|:  :| ::| :|:| |
Db        159 RQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYL 218

Qy        162 KVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI---------------------------- 195
              |:|      |  ||:||||||||||:|||| ||:
Db        219 KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLP 278

Qy        196 -----GH------NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGL 244
                  |:      :|:|||:|| : |  | ||||| || :  ||  ||| :   :|:||
Db        279 VLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGT 337

Qy        245 YPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAV 304
               | |||| ||||||:    : |  :::|||| | |: || ||::|| :| |:  |  :|
Db        338 GPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAA 397

Qy        305 RAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |: |    | |  |  :: || : |   ||  |:|| :|:|
Db        398 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 438


RESULT 15
US-10-200-910-8
; Sequence 8, Application US/10200910
; Publication No. US20030009021A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Hilbun, Erin
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20030009021A1el Human Proteases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0086-USA
;  CURRENT APPLICATION NUMBER: US/10/200,910
;  CURRENT FILING DATE:  2002-07-22
;  PRIOR APPLICATION NUMBER: US/09/710,099
;  PRIOR FILING DATE: 2000-11-10
;  PRIOR APPLICATION NUMBER: US 60/165,260
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 399
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-200-910-8

  Query Match             28.6%;  Score 546;  DB 15;  Length 399;
  Best Local Similarity   38.7%;  Pred. No. 5.3e-47;
  Matches  129;  Conservative   60;  Mismatches  124;  Indels   20;  Gaps   10;

Qy         23 FQSGQVLAALPRTSRQVQVLQNL--TTTYEIVLWQ-PVTADLIVKKKQVHFFVNASDVDN 79
              |   |||  | :  :|: :| :|      ::  |: |   | |  :     |  |:: :
Db         37 FTGDQVLRVLAKDEKQLSLLGDLEGLKPQKVDFWRGPARPSLPVDMR-----VPFSELKD 91

Qy         80 VKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEF 135
              :||:|   |:  |::: |:: |:   :| ::      |:: |: |  ||:| ||||||:
Db         92 IKAYLESHGLAYSIMIKDIQVLLDEERQAMAKSRRLERSTNSFSYSSYHTLEEIYSWIDN 151

Qy        136 ITERHPDMLTKIHIGSSFEKYPLYVLKVS--GKEQTAKNAIWIDCGIHAREWISPAFCLW 193
                | |:::|| ||:|||   : ||| |  |       ||||| |||:||||: |  :|
Db        152 FVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHP---AIWIDTGIHSREWITHATGIW 208

Qy        194 FIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVK 253
                ||:||||:|      ||| ||| |: |  :   |::|: ||||| |  |:|||||
Db        209 TANKNRLWRKNKSIRPGIFCIGVDLNRNWKS-GFGGNGSNSNPCSETYHGPSPQSEPEVA 267

Qy        254 AVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKN 313
              |: :|:  :  |  || ||:||||| :::||         :  ||  :| :|| |: |
Db        268 AIVNFITAHGN-FKALISIHSYSQMLMYPYGRLLEPVSNQRELYDLAKDAVEALYKV-HG 325

Qy        314 TRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |  |  | |||:| |   || || ||||:|:
Db        326 IEYIFGSISTTLYVASGITVDWAYDSGIKYAFS 358


Search completed: January  1, 2004, 19:50:56
Job time : 52.4823 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:25:56 ; Search time 24.8017 Seconds
(without alignments)
1395.903 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1799.5	94.2	423	2	A41204	carboxypeptidase B
2	692	36.2	415	2	A32129	carboxypeptidase B
3	628.5	32.9	417	1	A34487	carboxypeptidase A
4	621.5	32.5	416	1	A42332	carboxypeptidase B
5	619.5	32.4	417	2	A43929	carboxypeptidase A
6	580	30.4	306	1	CPBOB	carboxypeptidase B
7	553	28.9	417	2	A32128	carboxypeptidase A
8	546	28.6	309	2	A38395	mast cell carboxyp
9	533.5	27.9	419	1	CPBOA	carboxypeptidase A
10	528.5	27.7	419	1	CPRTA	carboxypeptidase A
11	508	26.6	417	2	A56171	carboxypeptidase A
12	489.5	25.6	419	1	S29127	carboxypeptidase A
13	437.5	22.9	528	2	T33527	hypothetical prote
14	397	20.8	323	2	T20507	hypothetical prote
15	395.5	20.7	667	2	T33526	hypothetical prote
16	391.5	20.5	527	2	T27572	hypothetical prote
17	369.5	19.3	303	1	CPCYB	carboxypeptidase B
18	362.5	19.0	455	2	T33572	hypothetical prote
19	358	18.7	666	2	T24170	hypothetical prote
20	299.5	15.7	430	2	S48976	ECM14 protein - ye
21	277.5	14.5	424	2	S17571	carboxypeptidase T
22	271	14.2	451	2	S20723	carboxypeptidase (
23	250	13.1	497	2	T40260	carboxypeptidase p
24	244.5	12.8	373	2	T26030	hypothetical prote
25	173	9.1	999	2	T36021	probable zinc-bind
26	135.5	7.1	1446	2	T30916	carboxypeptidase D
27	130.5	6.8	558	2	S75104	hypothetical prote
28	118.5	6.2	707	2	T40070	origin recognition
29	104.5	5.5	985	2	D82776	pyruvate dehydroge
30	103	5.4	1216	2	H84629	hypothetical prote
31	101.5	5.3	985	2	T29910	hypothetical prote
32	100	5.2	376	1	E69957	gamma-D-glutamyl-L
33	99.5	5.2	1389	2	I50090	carboxypeptidase g
34	99	5.2	1587	2	AB2012	hypothetical prote
35	98.5	5.2	983	2	B49284	immediate-early pr
36	98.5	5.2	1078	2	T44232	hypothetical prote
37	98	5.1	491	2	B96739	hypothetical prote
38	97	5.1	663	2	T37772	telomere length re
39	96	5.0	368	2	T21748	hypothetical prote
40	95.5	5.0	355	2	C83850	gamma-D-glutamyl-L
41	95	5.0	444	2	S51199	tryptophan 5-monoo
42	95	5.0	662	2	G86210	hypothetical prote
43	94	4.9	606	2	E90536	hypothetical prote
44	93.5	4.9	986	2	E90596	restriction-modifi
45	93.5	4.9	986	2	H90565	restriction modifi

ALIGNMENTS

RESULT 1
A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N;Alternate names: plasma carboxypeptidase B
C;Species: Homo sapiens (man)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C;Accession: A41204
R;Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A;Title: Isolation, molecular cloning, and partial characterization of a novel carboxype
A;Reference number: A41204; MUID:92042093; PMID:1939207
A;Accession: A41204
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <EAT>
A;Cross-references: GB:M75106; NID:g189686; PID:g189687
C;Genetics:
A;Gene: GDB:CPB2
A;Cross-references: GDB:129546; OMIM:212070
A;Map position: 13q14.11-13q14.11
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-114/Domain: activation peptide #status predicted <ACP>
F;115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Query Match 94.2%; Score 1799.5; DB 2; Length 423;
Best Local Similarity 89.6%; Pred. No. 3.2e-140;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

```
Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH----------------------------------NRMWRK 203
              |||||||||||||||||                                  ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
```

```
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C;Accession: A32129; S17543
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: A32129
A;Molecule type: DNA
A;Residues: 1-415 <CLA>
A;Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R;Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A;Title: Distribution of manganese in rat pancreas and identification of its primary bir
A;Reference number: S17543; MUID:91378950; PMID:1898371
A;Accession: S17543
A;Molecule type: protein
A;Residues: 109-116,'X',118-130 <KOD>
C;Genetics:
A;Gene: CPB
A;Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;354,376/Active site: Tyr, Glu #status predicted

  Query Match             36.2%;  Score 692;  DB 2;  Length 415;
  Best Local Similarity   39.2%;  Pred. No. 4e-49;
  Matches  150;  Conservative   61;  Mismatches  126;  Indels   46;  Gaps    7;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              | |    | | :    |:|   |   :|          | ::| |  | ||  |:| :|
Db          1 MLLLLALVSVALAHASEEH---FDGNRVYRVSVHGEDHVNLIQELANTKEIDFWKPDSAT 57

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
               :    | | | | || :|:  |  : :   ||:::| : :: |  : |   |||   |
Db         58 QVKPLTTVDFHVKAEDVADVENFLEENEVHYEVLISNVRNALESQFDSHT---RASGHSY 114

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              :|:    | :||: :   :||::|:  ||:::||   :||||: || :  | ||:||||
Db        115 TKYNKWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAIFIDCGF 173

Qy        181 HAREWISPAFCLWFI------------------------------GH------NRMWRK 203
              |||||||||| ||:                              |:      :|||||
Db        174 HAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRMWRK 233

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||  | : |:|   | || :  ||| ||| | |||||||  |||| | ||:| |:| |:
Db        234 TRSTMAGSSCLGVRPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNL 292

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              : ||||:::|||||  :::||||     :::|||: :   | : :  |   |:||:| |:
Db        293 STIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYTYGPGAT 351

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |:| | || ||| || |||||||
Db        352 TIYPAAGGSDDWSYDQGIKYSFT 374


RESULT 3
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34487
R;Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A;Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A;Reference number: A34487; MUID:90062123; PMID:2584208
A;Accession: A34487
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             32.9%;  Score 628.5;  DB 1;  Length 417;
  Best Local Similarity   37.0%;  Pred. No. 6.8e-44;
  Matches  140;  Conservative   60;  Mismatches  131;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKK 65
              :||:   :      |   |   :|      :  :   ||:||| : |:  | |      |
Db          7 MAVIYTTLAIAPVH---FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVN 63

Qy         66 KQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYH 124
              | | |:  :   ::: |    |   :|: |::: |::|    | :: | |   | :|:
Db         64 MTVDFRVSEKESQTIQSTLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
              ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::||||||||
Db        121 DWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy        185 WISPAFCLWFI-------GHN-----------------------------RMWRKNRSF 207
              ||||||| ||:       | |                             ||||||||
Db        180 WISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy        208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
              |: ||||||| ||    |     ::  |   | |  |||| | ||| :|:| ::| ||
Db        240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIK 298

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              |||: ||||| :: || ||     :|::|  ||  |  |:  |   ||| :|  : |:|
Db        299 AYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYK 357

Qy        328 APGGGDDWIYDLGIKYSF 345
              |   ||:||||||::|
Db        358 TSGSSLDWVYDLGIKHTF 375


RESULT 4
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N;Alternate names: carboxypeptidase B isozyme 1
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: A42332; S08254; S02813; S02812
R;Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992
A;Title: Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Ident
A;Reference number: A42332; MUID:92129345; PMID:1370825
A;Accession: A42332
A;Molecule type: mRNA
A;Residues: 1-416 <YAM>
A;Cross-references: GB:M81057; NID:g809194; PIDN:AAA66973.1; PID:g189625
A;Experimental source: pancreas
A;Note: sequence extracted from NCBI backbone (NCBIN:78573, NCBIP:78574)
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08254
A;Molecule type: protein
A;Residues: 16-40 <MOU>
A;Experimental source: pancreas
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
```

Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02813
A;Molecule type: protein
A;Residues: 16-36,'X',38-43 <PAS>
A;Experimental source: pancreas
A;Accession: S02812
A;Molecule type: protein
A;Residues: 16-36,'Q',38-43 <PA2>
A;Experimental source: pancreas
C;Genetics:
A;Gene: GDB:CPB1
A;Cross-references: GDB:132160; OMIM:114852
A;Map position: 3q21.3-3q25
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-110/Domain: activation peptide #status predicted <ACP>
F;111-415/Product: carboxypeptidase B, pancreatic #status predicted <MAT>
F;173-186,258-272/Disulfide bonds: #status predicted
F;176,179,303/Binding site: zinc (His, Glu, His) #status predicted
F;355,377/Active site: Tyr, Glu #status predicted

Query Match 32.5%; Score 621.5; DB 1; Length 416;
Best Local Similarity 35.1%; Pred. No. 2.5e-43;
Matches 134; Conservative 71; Mismatches 132; Indels 45; Gaps 8;

```
Qy          3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADL 61
              : :| ||| : |     |    |: :|          : ::: | :| :|  |:| :
Db          1 MLALLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQ 60

Qy         62 IVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYE 121
              |      | | | | |    |:  |  : :   ||:::: :::: |  :      ||:   ||
Db         61 IKPHSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYE 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|:    | :| : :   :| :::: ||::||   :|:||| ||    | ||::||| |
Db        118 KYNKWETIEAWTQQVATENPALISRSVIGTTPEGRAIYLLKV-GKAGQNKPAIFMDCGFH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              |||||||||| ||:                             |:      :| |||
Db        177 AREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKT 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              || :  :  |||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  ::
Db        237 RSTHTGS-SIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLS 294

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              ||||:::||||| :::||||     ::: ||: :|   |: :  :   |:||:| |: |
Db        295 SIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATT 353

Qy        325 LYLAPGGGDDWIYDLGIKYSFT 346
              :| | || ||| || ||:||||
Db        354 IYPAAGGSDDWAYDQGIRYSFT 375
```

RESULT 5
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N;Alternate names: mast cell carboxypeptidase A3
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: A43929; A39246; A45759
R;Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A;Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A
A;Reference number: A43929; MUID:92105393; PMID:1729276
A;Accession: A43929
A;Molecule type: DNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:M73716
A;Experimental source: mast cell
A;Note: the authors translated the codon CGC for residue 231 as Thr
R;Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafin,
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A;Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison
A;Reference number: A39246; MUID:90083291; PMID:2594780
A;Accession: A39246
A;Molecule type: mRNA
A;Residues: 1-417 <RE2>
A;Cross-references: GB:M27717; NID:g179933; PIDN:AAA35652.1; PID:g179934
R;Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.
J. Clin. Invest. 83, 1630-1636, 1989
A;Title: Human mast cell carboxypeptidase. Purification and characterization.
A;Reference number: A45759; MUID:89214692; PMID:2708524
A;Accession: A45759
A;Molecule type: protein
A;Residues: 110-137 <GOL>
C;Genetics:
A;Gene: GDB:CPA3
A;Cross-references: GDB:125231; OMIM:114851
A;Map position: 3q21.3-3q25
A;Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-109/Domain: activation peptide #status predicted <ACT>
F;110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F;176,179,304/Binding site: zinc (His, Glu, His) #status predicted
F;245-268/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Query Match 32.4%; Score 619.5; DB 2; Length 417;
Best Local Similarity 34.1%; Pred. No. 3.7e-43;
Matches 130; Conservative 76; Mismatches 128; Indels 47; Gaps 8;

```
Qy          6 LAVLVPIVLFCEQHVFA---FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLI 62
              : :::|: |       |    |   :|    |: :|  ::::|  | |:  | |     :
Db          1 MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHV 60

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYE 121
                 | | |:  :   ::: |: : :   :|: |::: |::|    : :  | |   |
Db         61 AANMMVDFRVSEKESQAIQSALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YA 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||||
Db        118 KYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              ||||:|||||  ||:                             |:      |||||||
Db        177 AREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKN 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||   |: ||||||| || :  |      ::  |:: | |  |||| | ||| :|:| ::|
Db        237 RSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              :|| ||: ||||| ::||| ||     :||:|: ||      :  |   ||| :|    |
Db        296 EIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIEST 354

Qy        325 LYLAPGGGDDWIYDLGIKYSF 345
              :|   |   || ||||||::|
Db        355 IYPISGSSLDWAYDLGIKHTF 375
```

RESULT 6
CPBOB
carboxypeptidase B (EC 3.4.17.2) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
C;Accession: A93797; A92150; A00912
R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A;Title: Amino-acid sequence of bovine carboxypeptidase B.
A;Reference number: A93797; MUID:75217824; PMID:1057162

```
A;Accession: A93797
A;Molecule type: protein
A;Residues: 1-306 <TIT>
R;Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations c
A;Reference number: A92150; MUID:74260705; PMID:4833744
A;Accession: A92150
A;Molecule type: protein
A;Residues: 31-93;131-181;263-265;292-306 <SCH>
R;Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A;Reference number: A92839; MUID:76265065; PMID:957425
A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R;Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptid
A;Reference number: A92051; MUID:70007159; PMID:5344132
A;Contents: annotation; active site
R;Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptida
A;Reference number: A92119; MUID:73061487; PMID:4565668
A;Contents: annotation; active site
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-158,149-163/Disfulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

  Query Match             30.4%;  Score 580;  DB 1;  Length 306;
  Best Local Similarity   43.6%;  Pred. No. 4.3e-40;
  Matches  115;  Conservative   44;  Mismatches   65;  Indels   40;  Gaps    5;

Qy        120 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 179
              ||:|::   | :| | :   :||::::  ||::|    :|:||| ||  : | |:::||| 
Db          6 YEKYNNWETIEAWTEQVASENPDLISRSAIGTTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy        180 IHAREWISPAFCLWFI------------------------------GH------NRMWR 202
              |||||||||||| ||:                             |:      |||||
Db         65 FHAREWISPAFCQWFVREAVRTYGREIHMTEFLDKLDFYVLPVVNIDGYIYTWTTNRMWR 124

Qy        203 KNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262
              | ||  | : | ||||| ||    ||  |||:: |||||||   ||| | |||| |:| :
Db        125 KTRSTRAGSSCTGTDLNRNF-DAGWCSIGASNNPCSETYCGSAAESEKESKAVADFIRNH 183

Qy        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322
              :: ||||:::||||| :::||||     |:: ||: :|  ||: :  :   | |::| |:
Db        184 LSSIKAYLTIHSYSQMMLYPYSYDYKLPKNNVELNTLAKGAVKKL-ASLHGTTYSYGPGA 242

Qy        323 ETLYLAPGGGDDWIYDLGIKYSFT 346
              |:| | || ||| || |||||||
Db        243 TTIYPASGGSDDWAYDQGIKYSFT 266


RESULT 7
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C;Accession: A32128
R;Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rut
J. Biol. Chem. 263, 17828-17836, 1988
A;Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and ev
A;Reference number: A32128; MUID:89034323; PMID:3182871
A;Accession: A32128
A;Molecule type: mRNA
A;Residues: 1-417 <GAR>
A;Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             28.9%;  Score 553;  DB 2;  Length 417;
  Best Local Similarity   33.5%;  Pred. No. 1.1e-37;
  Matches  137;  Conservative   69;  Mismatches  137;  Indels   66;  Gaps   13;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTT--YEIVLWQPVT 58
              |:|  |   :   ::|::    |   |||  :|    |:: |  |       |:  |:  |
Db          1 MRLTLLLAALLGYIYCQE---TFVGDQVLEIIPSHEEQIRTLLQLEAEEHLELDFWKSPT 57

Qy         59 ADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISN---DTVSPRA 115
                  :  : ||  |  : :  ||  |   ||  |::: ||: |: |:      :    |
Db         58 ----IPGETVHVRVPFASIQAVKVFLESQGIDYSIMIEDVQVLLDQEREEMLFNQQRERG 113

Qy        116 SASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAI 174
                :| ||:| |||  :: :   :| :::|:::|||||  |: ||| | | :   | ||
Db        114 GNFNPEAYHTLEEIYQEMDNLVAENPGLVSKVNLGSSFENRPMNVLKFSTGGD---KPAI 170

Qy        175 WIDCGIHAREWISPAFCLW-----------------------FI-------GH------ 197
              |:| |||||||:: |  ||                       |:       |:
Db        171 WLDAGIHAREWVTQATALWTANKIASDYGTDPAITSLLNTLDIFLLPVTNPDGYVFSQTT 230

Qy        198 NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 257
              ||||| ||  : : |:| | | |:   ::   ||||| ||::| |  | || |||::
Db        231 NRMWRKTRSKRSGSGCVGVDPNRNW-DANFGGPGASSSPCSDSYHGPKPNSEVEVKSIVD 289

Qy        258 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYT 317
              |:: :  ::||:|::||||| ::||| |  :|  |  ||  || :| :|: |    | |
Db        290 FIKSH-GKVKAFITLHSYSQLLMFPYGYKCTKPDDFNELDEVAQKAAQAL-KRLHGTSYK 347

Qy        318 HGHGSETLYLAPGGGDDWIYDLGIKYSFTSN----------PPVEKLLP 356
              |     :| | ||  || |||||||||              | :::||
Db        348 VGPICSVIYQASGGSIDWAYDLGIKYSFAFELRDTAFYGFLLPAKQILP 396


RESULT 8
A38395
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C;Accession: A38395; A33118
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A;Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme acti
A;Reference number: A38395; MUID:91105153; PMID:1988052
A;Accession: A38395
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-309 <COL>
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A;Reference number: A33118
A;Accession: A33118
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-309 <CO2>
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase
F;248,270/Active site: Tyr, Glu #status predicted

  Query Match             28.6%;  Score 546;  DB 2;  Length 309;
  Best Local Similarity   42.2%;  Pred. No. 2.8e-37;
  Matches  111;  Conservative   39;  Mismatches   73;  Indels   40;  Gaps    5;

Qy        120 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 179
              | :|:  |:| || | : |:||:|:::| |||: |  ||||||: |::   : ||::|||
Db          8 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 66

Qy        180 IHAREWISPAFCLWFI-------GHN-------------------------------RMWR 202
              ||||||:|||||  ||:       | |                               ||||
Db         67 IHAREWVSPAFCQWFVYQAAKSYGKNNIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 126

Qy        203 KNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262
              ||||   :: ||||||| ||    |     : : |   | |  |||| | ||| :|:| :
```

```
Db        127 KNRSKNPSSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 185

Qy        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322

Db        186 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy        323 ETLYLAPGGGDDWIYDLGIKYSF 345

Db        245 STIYKTSGSSLDWAYDLGIKHTF 267


RESULT 9
CPBOA
carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine
N;Alternate names: procarboxypeptidase A activation peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1992 #text_change 05-May-2000
C;Accession: JN0126; JT0440; A31406; S00059; A38834; A00910; A90355
R;Le Hueerou, I.; Guilloteau, P.; Toullec, R.; Puigserver, A.; Wicker, C.
Biochem. Biophys. Res. Commun. 175, 110-116, 1991
A;Title: Cloning and nucleotide sequence of a bovine pancreatic preprocarboxypeptidase A
A;Reference number: JN0126; MUID:91151335; PMID:1998496
A;Accession: JN0126
A;Molecule type: mRNA
A;Residues: 1-419 <LEH>
A;Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R;Foglizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A;Title: Primary structure of the activation peptide from bovine pancreatic procarboxype
A;Reference number: JT0440; MUID:88209583; PMID:3365436
A;Accession: JT0440
A;Molecule type: protein
A;Residues: 17-110 <FOG>
R;Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A;Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidas
A;Reference number: A31406; MUID:89150306; PMID:3147705
A;Accession: A31406
A;Molecule type: protein
A;Residues: 17-110 <WAD>
R;Chapus, C.; Kerfelec, B.; Foglizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A;Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A;Reference number: S00059; MUID:87275921; PMID:3609014
A;Accession: S00059
A;Molecule type: protein
A;Residues: 17-69 <CHA>
R;Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A;Title: The amino acid sequence of bovine carboxypeptidase A.
A;Reference number: A38834; MUID:70050818; PMID:5260942
A;Accession: A38834
A;Molecule type: protein
A;Residues: 111-137,'Q',139-140,'E',142-417 <BRA>
R;Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A;Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterizatic
A;Reference number: A90360; MUID:71155185; PMID:4927805
A;Accession: A00910
A;Molecule type: protein
A;Residues: 214-411 <BR2>
A;Note: this is the final paper in a series
R;Petra, P.H.; Hermodson, M.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 4023-4025, 1971
A;Title: Characterization of bovine carboxypeptidase A(Allan).
A;Reference number: A90355; MUID:72138789; PMID:5143102
A;Accession: A90355
A;Molecule type: protein
A;Residues: 111-142 <PET>
R;Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A;Title: Identification of the amino acid replacements characterizing the allotropic for
A;Reference number: A90556; MUID:69283620; PMID:5817619
A;Contents: annotation; allelic variant
A;Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CPA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137,'Q',139-1
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/Disulfide bonds: #status experimental
F;358,380/Active site: Tyr, Glu #status experimental

  Query Match           27.9%;  Score 533.5;  DB 1;  Length 419;
  Best Local Similarity  34.3%;  Pred. No. 4.5e-36;
  Matches  134;  Conservative   61;  Mismatches  133;  Indels   63;  Gaps   11;

Qy          3 LCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--W----QP 56

Db          4 LLILSVLLGAALGKED----FVGHQVLRITAADEAEVQTVKELEDLEHLQLDFWRGPGQP 59

Qy         57 VTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSP 113

Db         60 --------GSPIDVRVPFPSLQAVKVFLEAHGIRYRIMIEDVQSLLDEEQEQMFASQSRA 111

Qy        114 RASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKN 172

Db        112 RSTNTFNYATYHTLDEIYDFMDLLVAEHPQLVSKLQIGRSYEGRPIYVLKFS-TGGSNRP 170

Qy        173 AIWIDCGIHAREWISPAFCLWFI---------------------------------- 195

Db        171 AIWIDLGIHSREWITQATGVWFAKKFTEDYGQDPSFTAILDSMDIFLEIVTNPDGFAFTH 230

Qy        196 GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAV 255

Db        231 SQNRLWRKTRSVTSSSLCVGVDANRNW-DAGFGKAGASSSPCSETYHGKYANSEVEVKSI 289

Qy        256 ASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTR 315

Db        290 VDFVKDHGN-FKAFLSIHSYSQLLLYPYGYTTQSIPDKTELNQVAKSAVEAL-KSLYGTS 347

Qy        316 YTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346

Db        348 YKYGSIITTIYQASGGSIDWSYNQGIKYSFT 378


RESULT 10
CPRTA
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.L
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <QUI>
A;Cross-references: GB:J00713; NID:g203364; PIDN:AAA40893.1; PID:g203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA
```

```
A;Residues: 1-260,'F',262,'M',264-346,'K',348-419 <CLA>
C;Genetics:
A;Gene: CPA1
A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

  Query Match             27.7%;  Score 528.5;  DB 1;  Length 419;
  Best Local Similarity   34.4%;  Pred. No. 1.2e-35;
  Matches  135;  Conservative   61;  Mismatches  133;  Indels   63;  Gaps   11;

Qy          2 KLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTA 59
              :|  |::|:  |   |     |   |||         ||| :: |     : |  |:
Db          3 RLLILSLLLEAVCGNEN----FVGHQVLRISAADEAQVQKVKELEDLEHLQLDFWRDAAR 58

Qy         60 DLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRAS 116
                |   :| |      : :||| |   ||   ::: ||: |:   :||:|      ::
Db         59 AGIPIDVRVPF----PSIQSVKAFLEYHGISYEIMIEDVQLLLDEEKQQMSAFQARALST 114

Qy        117 ASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIW 175
               |: |  ||:|:||| ::: :   || :::|| ||::||  |::||| |    | : |||
Db        115 DSFNYATYHTLDEIYEFMDLLVAEHPQLVSKIQIGNTFEGRPIHVLKFS-TGGTNRPAIW 173

Qy        176 IDCGIHAREWISPAFCLWFIGH-----------------------------------N 198
              || |||:|||:: |  :||                                      |
Db        174 IDTGIHSREWVTQASGVWFAKKVTKDYGQDPTFTAVLDNMDIFLEIVTNPDGFAYTHKTN 233

Qy        199 RMWRKNRSFYANNHCIGTDLNSNFVSKHW----CEEGASSSSCSETYCGLYPESEPEVKA 254
              ||||| ||    : |:| | | |     |     : ||||: ||||| | :| || |||:
Db        234 RMWRKTRSHTQGSLCVGVDPNRN-----WDAGLGKAGASSNPCSETYRGKFPNSEVEVKS 288

Qy        255 VASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNT 314
              :  |:  : | |||:||:|||| :::|| ||   : |  ||  :|  || |:  :   |
Db        289 IVDFVTSHGN-IKAFISIHSYSQLLLYPYGYTSEPAPDQAELDQLAKSAVTAL-TSLHGT 346

Qy        315 RYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              : :|   :|:| | |   || |  ||||||
Db        347 EFKYGSIIDTIYQASGSTIDWTYSQGIKYSFT 378


RESULT 11
A56171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C;Accession: A56171; S02809; S71395
R;Catasus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAA74425.1; PID:g790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S',18-36,'N',38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LAE>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             26.6%;  Score 508;  DB 2;  Length 417;
  Best Local Similarity   32.5%;  Pred. No. 5.6e-34;
  Matches  128;  Conservative   63;  Mismatches  137;  Indels   66;  Gaps   11;

Qy         19 HVF---AFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTADLIVKKKQVHFFVN 73
               |::    |   |||  :|    |:: |  |      : |  |:  |       :  |  |
Db         13 HIYCLETFVGDQVLEIVPSNEEQIKNLLQLEAQEHLQLDFWKSPT----TPGETAHVRVP 68

Qy         74 ASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS---YYEQYHSLNEIY 130
                 :|  ||  |   ||  |::: ||: |: ::      : |   |    :  ||:| ||
Db         69 FVNVQAVKVFLESQGIAYSIMIEDVQVLLDKENEEMLFNRRRERSGNFNFGAYHTLEEIS 128

Qy        131 SWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIHAREWISPA 189
               :: :   || :::|::|||||  |: ||| | | :   | |||:| |||||||:: |
Db        129 QEMDNLVAEHPGLVSKVNIGSSFENRPMNVLKFSTGGD---KPAIWLDAGIHAREWVTQA 185

Qy        190 FCLWFIG-----------------------------------HNRMWRKNRSFYANNH 212
              ||                                       ||||||  ||  : :
Db        186 TALWTANKIVSDYGKDPSITSILDALDIFLLPVTNPDGYVFSQTKNRMWRKTRSKVSGSL 245

Qy        213 CIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISM 272
              |:| | | |:    :   ||||: ||::| |    || |||::  |:: :  ::||:| :
Db        246 CVGVDPNRNW-DAGFGGPGASSNPCSDSYHGPSANSEVEVKSIVDFIKSH-GKVKAFIIL 303

Qy        273 HSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGG 332
              ||||| ::||| |  :|  | :||| || :| ::: ::   |:|  |      :| | ||
Db        304 HSYSQLLMFPYGYKCTKLDDFDELSEVAQKAAQSL-RSLHGTKYKVGPICSVIYQASGGS 362

Qy        333 DDWIYDLGIKYSFTSN----------PPVEKLLP 356
              || || ||||||               |  ::||
Db        363 IDWSYDYGIKYSFAFELRDTGRYGFLLPARQILP 396


RESULT 12
S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N;Alternate names: pancreatic carboxypeptidase A1
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: S29127; A34205; S08253; S02810; S71394; S02811
R;Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver
Biochem. J. 287, 299-303, 1992
A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.
A;Reference number: S29127; MUID:93038569; PMID:1417781
A;Accession: S29127
A;Molecule type: mRNA
A;Residues: 1-419 <CAT>
A;Cross-references: EMBL:X67318; NID:g35329; PIDN:CAA47732.1; PID:g35330
R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A;Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
```

```
A;Molecule type: DNA
A;Residues: 330-396 <STE>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43;'XXX',114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

  Query Match             25.6%;  Score 489.5;  DB 1;  Length 419;
  Best Local Similarity   32.7%;  Pred. No. 1.9e-32;
  Matches  131;  Conservative   60;  Mismatches  127;  Indels   83;  Gaps   13;

Qy           3 LCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQ----- 55
               |  |:||:  |   |     |  |||          ||| :: |    :  |  |:
Db           4 LLVLSVLLGAVFGKED----FVGHQVLRISVADEAQVQKVKELEDLEHLQLDFWRGPAHP 59

Qy          56 --PVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQI---- 106
                 |:         :| |      :  ||  |   ||    :: ||: |:   |:|:
Db          60 GSPIDV-------RVPF----PSIQAVKIFLESHGISYETMIEDVQSLLDEEQEQMFAFR 108

Qy         107 ----SNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLK 162
                   | || :       |  ||:| ||| ::: :   :| :::|| ||:::|  |:|||
Db         109 SRARSTDTFN-------YATYHTLEEIYDFLDLLVAENPHLVSKIQIGNTYEGRPIYVLK 161

Qy         163 VSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------------------------- 195
                 |   : : ||||| |||:|||:: |  :||
Db         162 FS-TGGSKRPAIWIDTGIHSREWVTQASGVWFAKKITQDYGQDAAFTAILDTLDIFLEIV 220

Qy         196 ----------GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLY 245
                         |||||| ||  | : ||| | | |:    :   ||||: ||||| | :
Db         221 TNPDGFAFTHSTNRMWRKTRSHTAGSLCIGVDPNRNW-DAGFGLSGASSNPCSETYHGKF 279

Qy         246 PESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVR 305
                 || |||::  |:: : | |||:||:|||| :::|| |       | :||  ::  ||
Db         280 ANSEVEVKSIVDFVKDHGN-IKAFISIHSYSQLLMYPYGYKTEPVPDQDELDQLSKAAVT 338

Qy         306 AIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
               |:  :   |:: :|   : :| | |   || |  |||||||
Db         339 AL-ASLYGTKFNYGSIIKAIYQASGSTIDWTYSQGIKYSFT 378


RESULT 13
T33527
hypothetical protein T06A4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C;Accession: T33527
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33527
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-528 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.1
A;Map position: 1
A;Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C;Superfamily: carboxypeptidase

  Query Match             22.9%;  Score 437.5;  DB 2;  Length 528;
  Best Local Similarity   32.3%;  Pred. No. 4.8e-28;
  Matches  120;  Conservative   57;  Mismatches  122;  Indels   72;  Gaps   14;

Qy          33 PRTSRQVQVLQNL---TTTYEIVLWQPVT-----ADLIVKKKQVHFFVNASDVDNVKAHL 84
                 | |   |: |::|    : ||:  ||| |      || |       ||   |::: |
Db          54 PETEGSVKYLRSLYEDPSPYELDFWQPPTNIGAIVDLTVAPADAPRFV--KDLESKK--- 108

Qy          85 NVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDML 144
                   |   | : |:   |: :  :|      :   |::|:|| || : :: : : :| |:
Db         109 ----ISYIVAVNDLSKAIENERGSDKFYNPVAGFAYDKYNSLEEIQTEMKRLKKEYPTMI 164

Qy         145 TKIHIGSSFEKYPLYVLKVSGKEQT--AKNAIWIDCGIHAREWISPAFCLWFIGH----- 197
                | | || | |    | |:|::||     :| ::||| ||||||||:||  : :| |
Db         165 TLIDIGQSHENRTLLVMKITGKRNPLGSKISMWIDAGIHAREWIAPATAM-YIAHELLLG 223

Qy         198 ---------------------------------NRMWRKNRS-------FYANNHCIGTD 217
                                                |||||||||            ::   | | |
Db         224 YENDATVAKLMDHIDFYILPVMNPDGYEYSREKNRMWRKNRSPAKCARQTFSTVCCSGVD 283

Qy         218 LNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQ 277
                || ||     |   |:||  | :|| |     |||| :|| || :|  :  |:||:|||||
Db         284 LNRNF-DWFWASTGSSSDPCHDTYHGSAAFSEPESQAVRDFLEQNTPE--AFISLHSYSQ 340

Qy         278 HIVFPYSYTR-SKSKD-HEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDW 335
                 : || : :  |  :| |  |  :|  | :|: :    |:|  | |:: :| | ||  ||
Db         341 MWLIPYGHRKQSYPQDYHTGLRPLALRATKALYEL-YGTKYQVGTGADLMYEASGGSHDW 399

Qy         336 IY-DLGIKYSF 345
                  | : |::
Db         400 AKGQLKVPYAY 410


RESULT 14
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20507; T24289
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284
A;Accession: T20507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WIL>
A;Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone F02D8
R;Wild, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19869
A;Accession: T24289
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
A;Molecule type: DNA
A;Residues: 1-323 <WI2>
A;Cross-references: EMBL:Z78413; PIDN:CAB01666.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone T01C3
C;Genetics:
A;Gene: CESP:F02D8.4
A;Map position: 5
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C;Superfamily: carboxypeptidase

  Query Match            20.8%;  Score 397;  DB 2;  Length 323;
  Best Local Similarity   33.1%;  Pred. No. 5.3e-25;
  Matches   88;  Conservative   53;  Mismatches   81;  Indels   44;  Gaps    7;

Qy        132 WIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFC 191
              ::  : :::|: :   :||:|:|   :  :::: : ::|  :||| |||||||||
Db          3 FLNSLAQQYPNDVKLQNIGNSYEGRSITAVRIA-DDGSSKPIVWIDAGIHAREWISYNVA 61

Qy        192 LWFI--------------------------------GHNRMWRKNRSFYANNHCIGTDL 218
              |: |                               ::|||| || : |: | | |
Db         62 LYLIYTIVSQPAYRNLLDSVQLVVVPNTNPDGYEYSRTNDRMWRKTRSRFTNSRCAGADA 121

Qy        219 NSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQH 278
              | |:   :|  :| | | ||| :||  |:|||| |: : : |:  :|| ||::||| |
Db        122 NRNY-PFYWGTQGVSHSQCSEIFCGSRPQSEPEVLALTNAIIRDEERIKGYIALHSYGQE 180

Qy        279 IVFPYSYT-RSKSKDHEELSLVA---SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDD 334
              |::|: :| |:   | ::|  |    : |:||::   || ||  :  : || | |  || .
Db        181 ILYPWGHTQRTYPTDVQDLIQVGRAMASAIRAVN----NTDYTVVNSGDGLYPAAGASDD 236

Qy        335 WIYDLGIKYSFTSN-PPVEKLLPLSL 359
              |    |||||:|    |::     ||
Db        237 WAKSRGIKYSYTIELSPIDDFTGFSL 262


RESULT 15
T33526
hypothetical protein T06A4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33526
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-667 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67474.1; GSPDB:GN00019; CESP:T06A4.3
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.3
A;Map position: 1
A;Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

  Query Match            20.7%;  Score 395.5;  DB 2;  Length 667;
  Best Local Similarity   27.6%;  Pred. No. 1.9e-24;
  Matches  121;  Conservative   63;  Mismatches  138;  Indels  117;  Gaps   15;

Qy         12 IVLFCEQHV-FAPQSGQVLAALPRTSRQVQVLQNL---TTTYEIVLWQPVTA-----DLI 62
              .::  |  |  : ||  |   ||::   | :| |    | :::  |:         |::
Db          8 LIFLCFSAVSLSTQSFAVYRLLPKSQTDFQAIQRLYKNATDHDLNFWKTGKDKHGFWDVM 67

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVS---PR----- 114
              |  |   :|::          | |: |     : ||| || :  :| |:|   ||
Db         68 VDMKNSKYFLD---------FLQVNDISYIKTIDDVEGLITKHENNKTLSNMFPRLWDDS 118

Qy        115 ASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNA 173
              :|| | :  | |   :  |:: :  ::| |:  | || : |   :  ::: |  :| |
Db        119 SSAHYDFHTYGSYQRMTDWMKQLVVKYPKMVQYISIGKTTEGRNIDGVEIGGDSRT-KKI 177

Qy        174 IWIDCGIHAREWISPAFCLWFI---------------------------------- 195
              ||| ||||||| :|   |:||
Db        178 FWIDGGIHAREWAAPHTALFFIHQVCDPAPSEKSVFRLVPLGLHKCWQRQKKGKLDNKWL 237

Qy        196 ----------------------------------GHNRMWRKNRS--------FYAN 210
                                               | |:||||||        :  |
Db        238 TSRANEPGIKKLLNEITFVVVPCLNPDGYEFTRSSTNPHVRLWRKNRSKMQCRKDIWGRN 297

Qy        211 NHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFL--RRNINQIKA 268
              | | ||| ||   |: | | |   ||| | |  | |||| |||   |  :|   :  |
Db        298 RCCRGVDLNRNF-DFHFRESGTSDDPCSEIYQGPSPFSEPEAKAVRDALLSQRYKGRTDA 356

Qy        269 YISMHSYSQHIVFPYSYTR-SKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              ||::|:|||  : || : : :   | ::|  |  :| :|: |    |:|  | |::|||
Db        357 YITLHTYSQIWIHPYGHKKDAYPGDIKDLYEVGKKAAQAL-KRVYGTKYVVGSGADTLYP 415

Qy        328 APGGGDDWI-YDLGIKYSF 345
              | || :||  ::  :|: :
Db        416 ASGGSEDWAKHEAKVKFVY 434


Search completed: January  1, 2004, 19:34:29
Job time : 26.0517 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:03:36 ; Search time 14.2797 Seconds
(without alignments)
1185.568 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	692	36.2	415	1	CBPB_RAT	P19223 rattus norv
2	641	33.5	417	1	CBPB_HUMAN	P15086 homo sapien
3	629.5	32.9	401	1	CBPB_PIG	P09955 sus scrofa
4	628.5	32.9	417	1	CBPC_MOUSE	P15089 mus musculu
5	627	32.8	416	1	CBPB_CANFA	P55261 canis famil
6	619.5	32.4	417	1	CBPC_HUMAN	P15088 homo sapien
7	580	30.4	306	1	CBPB_BOVIN	P00732 bos taurus
8	553	28.9	417	1	CBP2_RAT	P19222 rattus norv
9	546	28.6	309	1	CBPC_RAT	P21961 rattus norv
10	533.5	27.9	419	1	CBPA_BOVIN	P00730 bos taurus
11	529.5	27.7	419	1	CBP1_RAT	P00731 rattus norv
12	511	26.7	417	1	CPB2_HUMAN	P48052 homo sapien
13	489.5	25.6	419	1	CBP1_HUMAN	P15085 homo sapien
14	483	25.3	421	1	CBP4_HUMAN	Q9ui42 homo sapien
15	404.5	21.2	304	1	CBPZ_SIMVI	P42788 simulium vi
16	386	20.2	433	1	CBPA_ANOGA	O02350 anopheles g
17	369.5	19.3	303	1	CBPB_ASTFL	P04069 astacus flu
18	299.5	15.7	430	1	YHT2_YEAST	P38836 saccharomyc
19	277.5	14.5	424	1	CBPT_THEVU	P29068 thermoactin
20	271	14.2	451	1	CBPS_STRGR	P18143 streptomyce
21	251	13.1	434	1	CBPS_STRCP	P39041 streptomyce
22	118.5	6.2	707	1	ORC1_SCHPO	P54789 schizosacch
23	114.5	6.0	1380	1	CBPD_HUMAN	O75976 homo sapien
24	112.5	5.9	1377	1	CBPD_MOUSE	O89001 mus musculu
25	105	5.5	1255	1	PER2_HUMAN	O15055 homo sapien
26	103.5	5.4	722	1	CPXM_MOUSE	Q9z100 mus musculu
27	102	5.3	734	1	CPXM_HUMAN	Q96sm3 homo sapien
28	100	5.2	376	1	YQGT_BACSU	P54497 bacillus su
29	99.5	5.2	445	1	TPH1_CHICK	P70080 gallus gall
30	98.5	5.2	488	1	TY3H_ANGAN	O42091 anguilla an
31	97	5.1	663	1	TAZ1_SCHPO	P79005 schizosacch
32	95	5.0	444	1	TPH1_HUMAN	P17752 homo sapien
33	91.5	4.8	465	1	STHA_ECO57	Q8x727 escherichia
34	91.5	4.8	465	1	STHA_ECOLI	P27306 escherichia
35	91.5	4.8	3588	1	SRF1_BACSU	P27206 bacillus su
36	91	4.8	297	1	Y103_HUMAN	Q15006 homo sapien
37	91	4.8	444	1	TPH1_RABIT	P17290 oryctolagus
38	91	4.8	982	1	ENV_SFV3L	P27399 simian foam
39	90	4.7	1451	1	SPT6_YEAST	P23615 saccharomyc
40	89.5	4.7	682	1	CJBA_BACUH	O86170 bacillus th
41	89	4.7	525	1	PUR9_BUCAI	P57143 b bifunctio
42	88.5	4.6	319	1	UN30_CAEEL	P52906 caenorhabdi
43	88.5	4.6	466	1	STHA_SALTY	Q8xfp5 salmonella
44	88	4.6	454	1	CBPH_LOPAM	P37892 lophius ame
45	88	4.6	1150	1	C9EA_BACTA	Q9znl9 bacillus th

ALIGNMENTS

RESULT 1
CBPB_RAT
ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B genes. Comparative analysis of the rat carboxypeptidase gene family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O = peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration CC between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M23959; AAA40872.1; -.
DR EMBL; M23947; AAA40872.1; JOINED.
DR EMBL; M23950; AAA40872.1; JOINED.
DR EMBL; M23952; AAA40872.1; JOINED.
DR EMBL; M23953; AAA40872.1; JOINED.
DR EMBL; M23954; AAA40872.1; JOINED.
DR PIR; A32129; A32129.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.

FT SIGNAL	1	13	BY SIMILARITY.
FT PROPEP	14	108	ACTIVATION PEPTIDE.
FT CHAIN	109	415	CARBOXYPEPTIDASE B.
FT METAL	174	174	ZINC (BY SIMILARITY).
FT METAL	177	177	ZINC (BY SIMILARITY).
FT METAL	302	302	ZINC (BY SIMILARITY).

```
FT   ACT_SITE    376    376       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    171    184       BY SIMILARITY.
FT   DISULFID    243    266       BY SIMILARITY.
FT   DISULFID    257    271       BY SIMILARITY.
SQ   SEQUENCE   415 AA;  47515 MW;  8EA06CCADE30B6F2 CRC64;

  Query Match           36.2%;  Score 692;  DB 1;  Length 415;
  Best Local Similarity  39.2%;  Pred. No. 4.1e-51;
  Matches  150;  Conservative   61;  Mismatches  126;  Indels   46;  Gaps    7;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              | |    | | :    |:|   |   :|        | ::| |  | ||  |:| :|
Db          1 MLLLLALVSVALAHASEEH---FDGNRVYRVSVHGEDHVNLIQELANTKEIDFWKPDSAT 57

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
               :     | | | | || :|:  |  : :   ||:::| : :: |  : |   |||   |
Db         58 QVKPLTTVDFHVKAEDVADVENFLEENEVHYEVLISNVRNALESQFDSHT---RASGHSY 114

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              :|:    | :||: :   :||::|:  ||::||   :|||: || :  | ||:||||
Db        115 TKYNKWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAIFIDCGF 173

Qy        181 HAREWISPAFCLWFI----------------------------GH------NRMWRK 203
              |||||||||| ||:                            |:      :|||||
Db        174 HAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRMWRK 233

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
               ||  | : |:|   | || :  ||| ||| | |||||||  |||| | ||:| |:| |:
Db        234 TRSTMAGSSCLGVRPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNL 292

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              : ||||:::||||| :::|||      :::|||: :   | : :  |   |:||:| |:
Db        293 STIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYTYGPGAT 351

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |:| | || ||| || |||||||
Db        352 TIYPAAGGSDDWSYDQGIKYSFT 374


RESULT 2
CBPB_HUMAN
ID   CBPB_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15086; O60834; Q96BQ8;
DT   01-APR-1990 (Rel. 14, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE   (PASP).
GN   CPB1 OR CPB OR PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC   TISSUE=Pancreas;
RX   MEDLINE=92129345; PubMed=1370825;
RA   Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA   French C.K.;
RT   "Isolation of a cDNA encoding a human serum marker for acute
RT   pancreatitis. Identification of pancreas-specific protein as
RT   pancreatic procarboxypeptidase B.";
RL   J. Biol. Chem. 267:2575-2581(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=98182241; PubMed=9524066;
RA   Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA   Aviles F.X.;
RT   "Comparative analysis of the sequences and three-dimensional models
RT   of human procarboxypeptidases A1, A2 and B.";
RL   Biol. Chem. 379:149-155(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 16-43.
RC   TISSUE=Pancreas;
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- TISSUE SPECIFICITY: Pancreas.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81057; AAA66973.1; -.
DR   EMBL; AJ224866; CAA12163.1; -.
DR   EMBL; BC015338; AAH15338.1; -.
DR   PDB; 1KWM; 28-AUG-02.
DR   MEROPS; M14.003; -.
DR   Genew; HGNC:2299; CPB1.
DR   MIM; 114852; -.
DR   GO; GO:0004180; F:carboxypeptidase activity; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL        1     15
FT   PROPEP       16    110       ACTIVATION PEPTIDE.
FT   CHAIN       111    417       CARBOXYPEPTIDASE B.
FT   METAL       176    176       ZINC (BY SIMILARITY).
FT   METAL       179    179       ZINC (BY SIMILARITY).
FT   METAL       304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE    378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    173    186       BY SIMILARITY.
FT   DISULFID    245    268       BY SIMILARITY.
FT   DISULFID    259    273       BY SIMILARITY.
FT   CONFLICT     16     16       H -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     17     17       H -> Q (IN REF. 1; AA SEQUENCE).
```

```
FT   CONFLICT     37     37       H -> Q (IN REF. 4).
FT   CONFLICT    208    208       D -> N (IN REF. 2).
FT   CONFLICT    245    245       MISSING (IN REF. 1).
SQ   SEQUENCE   417 AA;  47367 MW;  EBBB98B27F5D5AF9 CRC64;

  Query Match           33.5%;  Score 641;  DB 1;  Length 417;
  Best Local Similarity   35.3%;  Pred. No. 8.6e-47;
  Matches  135;  Conservative   71;  Mismatches  132;  Indels   44;  Gaps    7;

Qy          3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADL 61
              : :| ||| : |     |    |:  :|          : ::: | :| :|  |:| :
Db          1 MLALLVLVTVALASAHHGGEHFEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQ 60

Qy         62 IVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYE 121
              |     | | | | |   |:  |  : :   ||:::: :::: | :     ||:   ||
Db         61 IKPHSTVDFRVKAEDTVTVENVLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYE 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||::||| |
Db        118 KYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              |||||||||| ||:                              |:      :| |||
Db        177 AREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKT 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              || :  : |||| | ||    ||| ||| : | |||||   ||| | ||:| |:|  ::
Db        237 RSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLS 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
               ||||:::||||| :::|||      ::: ||: :|   |: :  :   |:||:| |: |
Db        296 SIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATT 354

Qy        325 LYLAPGGGDDWIYDLGIKYSFT 346
              :| | || ||| || ||:||||
Db        355 IYPAAGGSDDWAYDQGIRYSFT 376


RESULT 3
CBPB_PIG
ID   CBPB_PIG       STANDARD;      PRT;   401 AA.
AC   P09955;
DT   01-MAR-1989 (Rel. 10, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2).
GN   CPB.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE OF 1-103.
RX   MEDLINE=91208150; PubMed=2018774;
RA   Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA   Aviles F.X.;
RT   "Analysis of the activation process of porcine procarboxypeptidase B
RT   and determination of the sequence of its activation segment.";
RL   Biochemistry 30:4082-4089(1991).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 1-38.
RX   MEDLINE=85279427; PubMed=4026847;
RA   Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT   "Sequential homologies between procarboxypeptidases A and B from
RT   porcine pancreas.";
RL   Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=91114690; PubMed=1989878;
RA   Coll M., Guasch A., Aviles F.X., Huber R.;
RT   "Three-dimensional structure of porcine procarboxypeptidase B: a
RT   structural basis of its inactivity.";
RL   EMBO J. 10:1-9(1991).
RN   [4]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX   MEDLINE=91027767; PubMed=2223783;
RA   Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT   "Sequence-specific 1H NMR assignments and determination of the
RT   secondary structure for the activation domain isolated from
RT   pancreatic procarboxypeptidase B.";
RL   Biochemistry 29:7515-7522(1990).
RN   [5]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=91114693; PubMed=1989879;
RA   Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
RT   "The NMR structure of the activation domain isolated from porcine
RT   procarboxypeptidase B.";
RL   EMBO J. 10:11-15(1991).
RN   [6]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=93044373; PubMed=1422143;
RA   Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA   Huber R., Wuethrich K.;
RT   "Comparison of the NMR solution structure with the X-ray crystal
RT   structure of the activation domain from procarboxypeptidase B.";
RL   J. Biomol. NMR 2:1-10(1992).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR   PDB; 1PBA; 31-OCT-93.
DR   PDB; 1NSA; 24-DEC-97.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW   3D-structure.
FT   PROPEP        1     95       ACTIVATION PEPTIDE.
FT   CHAIN        96    401       CARBOXYPEPTIDASE B.
FT   DISULFID    158    171
FT   DISULFID    230    253
FT   DISULFID    244    258
FT   METAL       161    161       ZINC.
FT   METAL       164    164       ZINC.
FT   METAL       289    289       ZINC.
FT   ACT_SITE    341    341
FT   ACT_SITE    363    363       NUCLEOPHILE.
FT   TURN          8      9
FT   STRAND       11     17
FT   HELIX        20     32
FT   STRAND       36     39
FT   HELIX        43     45
FT   STRAND       50     56
FT   HELIX        58     60
FT   HELIX        61     69
FT   TURN         70     72
FT   STRAND       74     79
FT   HELIX        82     87
FT   TURN         88     88
FT   TURN        101    102
FT   HELIX       107    120
FT   TURN        122    124
FT   STRAND      125    132
FT   TURN        134    135
FT   STRAND      138    144
FT   TURN        149    150
FT   STRAND      153    157
FT   TURN        162    163
FT   HELIX       165    181
```

```
FT   TURN        182    184
FT   HELIX       186    194
FT   STRAND      196    200
FT   HELIX       205    213
FT   TURN        214    214
FT   TURN        216    217
FT   STRAND      224    224
FT   TURN        226    227
FT   STRAND      231    231
FT   HELIX       235    237
FT   TURN        243    244
FT   STRAND      249    249
FT   TURN        252    253
FT   TURN        255    256
FT   STRAND      257    257
FT   TURN        262    263
FT   HELIX       266    277
FT   TURN        278    281
FT   STRAND      282    289
FT   STRAND      294    297
FT   TURN        307    308
FT   HELIX       309    327
FT   STRAND      332    335
FT   HELIX       336    339
FT   HELIX       347    353
FT   TURN        354    355
FT   STRAND      359    363
FT   TURN        371    372
FT   HELIX       376    378
FT   HELIX       379    398
FT   TURN        399    401
SQ   SEQUENCE   401 AA;  45713 MW;  53129AF159A26348 CRC64;

  Query Match             32.9%;  Score 629.5;  DB 1;  Length 401;
  Best Local Similarity   37.1%;  Pred. No. 7.7e-46;
  Matches  134;  Conservative   64;  Mismatches  120;  Indels   43;  Gaps    6;

Qy          23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
               |:  :|          :  |  | :| :|  |:| :   |     | | | | |:  |:
Db           7 FEGEKVFRVNVEDENDISELHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 66

Qy          83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
                 |  : :   ||: ::  ::: |  :     | :   ||:|::   | :| | :| ::||
Db          67 FLEQNELQYEVLINNLRSVLEAQFDSRC---RTTGHSYEKYNNWETIEAWTEQVTSKNPD 123

Qy         143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
               ::::  ||::|:   :|:||| ||  : | ||::||| ||||||| ||| ||:
Db         124 LISRSAIGTTFDGDNIYLLKV-GKPGSNKPAIFMDCGFHAREWISQAFCQWFVRDAVRTY 182

Qy         196 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                      |:      ||||||  ||  | : | ||| | || :
Db         183 GYEAHMTEFLDNLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGSSCTGTDPNRNF-NA 241

Qy         226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
                ||  ||| : |:|||||   ||| | ||:| |:|  |:: ||||:::|||||  |::||||
Db         242 GWCTVGASVNPCNETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 301

Qy         286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
                     ::  ||: :|  ||: :  :    | |::| || |:| | || ||| |: ||||||
Db         302 DYKLPENDAELNSLAKGAVKEL-ASLYGTSYSYGPGSTTIYPAAGGSDDWAYNQGIKYSF 360

Qy         346 T 346
               |
Db         361 T 361


RESULT 4
CBPC_MOUSE
ID   CBPC_MOUSE     STANDARD;      PRT;   417 AA.
AC   P15089;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90062123; PubMed=2584208;
RA   Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA   Serafin W.E.;
RT   "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT   novel member of the carboxypeptidase gene family.";
RL   J. Biol. Chem. 264:20094-20099(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05118; AAA37369.1; -.
DR   PIR; A34487; A34487.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   MGD; MGI:88479; Cpa3.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     15
FT   PROPEP       16    109       ACTIVATION PEPTIDE.
FT   CHAIN       110    417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL       176    176       ZINC (BY SIMILARITY).
FT   METAL       179    179       ZINC (BY SIMILARITY).
FT   METAL       304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE    378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    173    186       BY SIMILARITY.
FT   DISULFID    245    268       BY SIMILARITY.
SQ   SEQUENCE   417 AA;  48790 MW;  A2B300A068D1BA6D CRC64;

  Query Match             32.9%;  Score 628.5;  DB 1;  Length 417;
  Best Local Similarity   37.0%;  Pred. No. 9.8e-46;
  Matches  140;  Conservative   60;  Mismatches  131;  Indels   47;  Gaps    8;

Qy           6 LAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKK 65
               :||:   :      |   |   :|      :  :   ||:||| : |:  | |      |
Db           7 MAVIYTTLAIAPVH---FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVN 63

Qy          66 KQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYH 124
                 | | |:  :   ::: |     |   :|: |::: |::|     | :: | |    | :|:
Db          64 MTVDFRVSEKESQTIQSTLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYN 120

Qy         125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184
                 ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::||||||||
Db         121 DWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHARE 179

Qy         185 WISPAFCLWFI-------GHN-----------------------------RMWRKNRSF 207
               |||||||  ||:       | |                             ||||||||
Db         180 WISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSR 239

Qy         208 YANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIK 267
```

```
              |: ||||||| ||    |     ::  |   | |  |||| | ||| :|:| ::| ||
Db        240 NQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIK 298

Qy        268 AYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYL 327
              |||: ||||| :: || ||     :|::|  ||  |  |:  |   ||| :|  : |:|
Db        299 AYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYK 357

Qy        328 APGGGDDWIYDLGIKYSF 345
                |   ||:|||||::|
Db        358 TSGSSLDWVYDLGIKHTF 375


RESULT 5
CBPB_CANFA
ID   CBPB_CANFA     STANDARD;      PRT;   416 AA.
AC   P55261;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE   membrane associated protein) (ZAP47).
GN   CPB1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Fukuoka S.-I.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D78348; BAA11366.1; -.
DR   HSSP; P09955; 1PBA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     15       BY SIMILARITY.
FT   PROPEP       16    109       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN       110    416       CARBOXYPEPTIDASE B.
FT   METAL       175    175       ZINC (BY SIMILARITY).
FT   METAL       178    178       ZINC (BY SIMILARITY).
FT   METAL       303    303       ZINC (BY SIMILARITY).
FT   ACT_SITE    377    377       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    244    267       BY SIMILARITY.
FT   DISULFID    258    272       BY SIMILARITY.
SQ   SEQUENCE   416 AA;  47595 MW;  DFE1992CD52F8AB4 CRC64;

  Query Match             32.8%;  Score 627;  DB 1;  Length 416;
  Best Local Similarity   35.5%;  Pred. No. 1.3e-45;
  Matches  135;  Conservative   68;  Mismatches  127;  Indels   50;  Gaps    8;

Qy          8 VLVPIVL----FCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIV 63
              :|| : |    :  :|   |:  :|          : :|  | :| :|  |:| :   |
Db          5 ILVTLALASAHYSGEH---FEGEKVFRVNVEDENHINLLHTLASTTQIDFWKPDSVTQIK 61

Qy         64 KKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQY 123
                  | | | |:  |:  |  : :   ||: ::  ::: |          |:   ||:|
Db         62 PHSTADFRVKAEDILTVEDFLKQNELHYEVLINNLRLVLEGQFGRQV---PATGHSYEKY 118

Qy        124 HSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAR 183
              :    | :| : :|  :||:::  ||::||   :|:||| ||    | ||::||| |||
Db        119 NRWETIEAWTQQVTSENPDLISRRSIGTTFEGRTIYLLKV-GKAGQNKPAIFMDCGFHAR 177

Qy        184 EWISPAFCLWFI------------------------------GH------NRMWRKNRS 206
              |||||||  ||:                             |:      ||||| ||
Db        178 EWISPAFWQWFVREXIRTYGQEIHMTELLDKLDFYVLPVGNIDGYVYTWTKNRMWRKTRS 237

Qy        207 FYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQI 266
                :|:|||   ||    ||: ||| : | |||||   ||| | ||:|:|:| |:: |
Db        238 TQVGTNCVGTDPTRNF-DAGWCKIGASRNPCDETYCGPAAESEKETKALANFIRSNLSSI 296

Qy        267 KAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLY 326
              |||:::||||| :::||||    :::: ||: :|   |: :  |   |:||:| |: |:|
Db        297 KAYLTIHSYSQMMLYPYSYDYKLTENNAELNALAKATVKEL-ATLHGTKYTYGPGATTIY 355

Qy        327 LAPGGGDDWIYDLGIKYSFT 346
              | || ||| || |||||||
Db        356 PAAGGSDDWAYDQGIKYSFT 375


RESULT 6
CBPC_HUMAN
ID   CBPC_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15088; Q96E94;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=90083291; PubMed=2594780;
RA   Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA   Austen K.F., Serafin W.E.;
RT   "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT   comparison of the protein with mouse mast cell carboxypeptidase A and
RT   rat pancreatic carboxypeptidases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mast cells;
RX   MEDLINE=92105393; PubMed=1729276;
RA   Reynolds D.S., Gurley D.S., Austen K.F.;
RT   "Cloning and characterization of the novel gene for mast cell
RT   carboxypeptidase A.";
RL   J. Clin. Invest. 89:273-282(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

```
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 110-417 FROM N.A.
RX   MEDLINE=92333165; PubMed=1629626;
RA   Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,
RA   Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT   "Human skin mast cell carboxypeptidase: functional characterization,
RT   cDNA cloning, and genealogy.";
RL   J. Invest. Dermatol. 99:138-145(1992).
RN   [5]
RP   SEQUENCE OF 110-137.
RX   MEDLINE=89214692; PubMed=2708524;
RA   Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT   "Human mast cell carboxypeptidase. Purification and
RT   characterization.";
RL   J. Clin. Invest. 83:1630-1636(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27717; AAA35652.1; -.
DR   EMBL; M73720; AAA59568.1; -.
DR   EMBL; M73716; AAA59568.1; JOINED.
DR   EMBL; M73717; AAA59568.1; JOINED.
DR   EMBL; M73718; AAA59568.1; JOINED.
DR   EMBL; M73719; AAA59568.1; JOINED.
DR   EMBL; BC012613; AAH12613.1; -.
DR   EMBL; S40234; AAB22578.2; ALT_SEQ.
DR   PIR; A43929; A43929.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   Genew; HGNC:2298; CPA3.
DR   MIM; 114851; -.
DR   GO; GO:0030141; C:secretory granule; NAS.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    15
FT   PROPEP      16   109       ACTIVATION PEPTIDE.
FT   CHAIN      110   417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL      176   176       ZINC (BY SIMILARITY).
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173   186       BY SIMILARITY.
FT   DISULFID   245   268       BY SIMILARITY.
FT   CONFLICT    63    63       N -> K (IN REF. 3).
SQ   SEQUENCE   417 AA;  48700 MW;  8CB90DB758117B24 CRC64;
```

```
  Query Match           32.4%;  Score 619.5;  DB 1;  Length 417;
  Best Local Similarity   34.1%;  Pred. No. 5.7e-45;
  Matches  130;  Conservative   76;  Mismatches  128;  Indels   47;  Gaps    8;

Qy          6 LAVLVPIVLFCEQHVFA---FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLI 62
              : :::|: |     |   |  :|    |: :|  ::::|  | |:  | |     :
Db          1 MRLILPVGLIATTLAIAPVRFDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHV 60

Qy         63 VKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYE 121
                | | |:  :   ::: |: : :   :|: |::: |::|    : :  | |   |
Db         61 AANMMVDFRVSEKESQAIQSALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YA 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||||
Db        118 KYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIH 176

Qy        182 AREWISPAFCLWFI-----------------------------GH------NRMWRKN 204
              ||||:||||| ||:                             |:      ||||||
Db        177 AREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKN 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||   |: ||||||| || :  |     ::  |:: | |  |||| | ||| :|:| ::|
Db        237 RSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLN 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
              :|| ||: ||||| ::||| ||     :||:|: ||      :  |   ||| :|    |
Db        296 EIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIEST 354

Qy        325 LYLAPGGGDDWIYDLGIKYSF 345
              :|   |   || ||||||::|
Db        355 IYPISGSSLDWAYDLGIKHTF 375
```

```
RESULT 7
CBPB_BOVIN
ID   CBPB_BOVIN     STANDARD;      PRT;   306 AA.
AC   P00732;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B (EC 3.4.17.2).
GN   CPB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=75217824; PubMed=1057162;
RA   Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT   "Amino-acid sequence of bovine carboxypeptidase B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN   [2]
RP   SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX   MEDLINE=74260705; PubMed=4833744;
RA   Schmidt J.J., Hirs C.H.W.;
RT   "Primary structure of bovine carboxypeptidase B. Inferences from the
RT   locations of the half-cystines and identification of the active site
RT   arginine.";
RL   J. Biol. Chem. 249:3756-3764(1974).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX   MEDLINE=76265065; PubMed=957425;
RA   Schmid M.F., Herriott J.R.;
RT   "Structure of carboxypeptidase B at 2.8-A resolution.";
RL   J. Mol. Biol. 103:175-190(1976).
RN   [4]
RP   ACTIVE SITE.
RX   MEDLINE=70007159; PubMed=5344132;
RA   Plummer T.H. Jr.;
RT   "Isolation and sequence of peptides at the active center of bovine
RT   carboxypeptidase B.";
```

```
RL   J. Biol. Chem. 244:5246-5253(1969).
RN   [5]
RP   ACTIVE SITE.
RX   MEDLINE=73061487; PubMed=4565668;
RA   Kimmel M.T., Plummer T.H. Jr.;
RT   "Identification of a glutamic acid at the active center of bovine
RT   carboxypeptidase B.";
RL   J. Biol. Chem. 247:7864-7869(1972).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A93797; CPBOB.
DR   PDB; 1CPB; 30-SEP-83.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT   DISULFID     63     76
FT   DISULFID    135    158
FT   DISULFID    149    163
FT   METAL        66     66       ZINC.
FT   METAL        69     69       ZINC.
FT   METAL       194    194       ZINC.
FT   ACT_SITE    246    246
FT   ACT_SITE    268    268       NUCLEOPHILE.
SQ   SEQUENCE   306 AA;  34612 MW;  C329D2655C44A172 CRC64;

  Query Match             30.4%;  Score 580;  DB 1;  Length 306;
  Best Local Similarity   43.6%;  Pred. No. 8.2e-42;
  Matches  115;  Conservative   44;  Mismatches   65;  Indels   40;  Gaps    5;

Qy        120 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 179
              ||:|::   | :| | :   :||::::  ||::|    :|:||| ||  : | |:::|||
Db          6 YEKYNNWETIEAWTEQVASENPDLISRSAIGTTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy        180 IHAREWISPAFCLWFI------------------------------GH------NRMWR 202
              |||||||||||| ||:                              |:      |||||
Db         65 FHAREWISPAFCQWFVREAVRTYGREIHMTEFLDKLDFYVLPVVNIDGYIYTWTTNRMWR 124

Qy        203 KNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262
              | ||  | : | ||||| ||    ||  |||:: |||||||   ||| | |||| |:| :
Db        125 KTRSTRAGSSCTGTDLNRNF-DAGWCSIGASNNPCSETYCGSAAESEKESKAVADFIRNH 183

Qy        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322
              :: ||||:::||||| :::||||     |:: ||: :|  ||:':  :   | |::| |:
Db        184 LSSIKAYLTIHSYSQMMLYPYSYDYKLPKNNVELNTLAKGAVKKL-ASLHGTTYSYGPGA 242

Qy        323 ETLYLAPGGGDDWIYDLGIKYSFT 346
              |:| | || ||| || |||||||
Db        243 TTIYPASGGSDDWAYDQGIKYSFT 266


RESULT 8
CBP2_RAT
ID   CBP2_RAT       STANDARD;      PRT;   417 AA.
AC   P19222;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=89034323; PubMed=3182871;
RA   Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA   Graf M., Rutter W.J.;
RT   "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT   cloning, and evolutionary implications on substrate specificity in
RT   the carboxypeptidase gene family.";
RL   J. Biol. Chem. 263:17828-17836(1988).
RN   [2]
RP   SEQUENCE OF 131-143 FROM N.A.
RX   MEDLINE=95386501; PubMed=7657630;
RA   Normant E., Gros C., Schwartz J.C.;
RT   "Carboxypeptidase A isoforms produced by distinct genes or
RT   alternative splicing in brain and other extrapancreatic tissues.";
RL   J. Biol. Chem. 270:20543-20549(1995).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=92105124; PubMed=1761558;
RA   Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT   "Structural evolution of an enzyme specificity. The structure of rat
RT   carboxypeptidase A2 at 1.9-A resolution.";
RL   J. Biol. Chem. 266:24606-24612(1991).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23721; AAA40956.1; -.
DR   EMBL; M23714; AAA40956.1; JOINED.
DR   EMBL; M23715; AAA40956.1; JOINED.
DR   EMBL; M23716; AAA40956.1; JOINED.
DR   EMBL; M23717; AAA40956.1; JOINED.
DR   EMBL; M23718; AAA40956.1; JOINED.
DR   EMBL; M23719; AAA40956.1; JOINED.
DR   EMBL; M23720; AAA40956.1; JOINED.
DR   EMBL; S79837; -; NOT_ANNOTATED_CDS.
DR   PIR; A32128; A32128.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.002; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16       POTENTIAL.
FT   PROPEP      17    112       ACTIVATION PEPTIDE.
FT   CHAIN      113    417       CARBOXYPEPTIDASE A2.
FT   METAL      177    177       ZINC.
FT   METAL      180    180       ZINC.
FT   METAL      304    304       ZINC.
FT   ACT_SITE   378    378       NUCLEOPHILE.
FT   DISULFID   246    269       BY SIMILARITY.
FT   DISULFID   318    352       PROBABLE.
SQ   SEQUENCE   417 AA;  46912 MW;  BEDCC41A830F2D45 CRC64;

  Query Match             28.9%;  Score 553;  DB 1;  Length 417;
  Best Local Similarity   33.5%;  Pred. No. 2.4e-39;
  Matches  137;  Conservative   69;  Mismatches  137;  Indels   66;  Gaps   13;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTT--YEIVLWQPVT 58
              |:|  |   :   ::|::    |   |||  :|    |:: |  |      |:  |:  |
Db          1 MRLTLLLAALLGYIYCQE---TFVGDQVLEIIPSHEEQIRTLLQLEAEEHLELDFWKSPT 57

Qy         59 ADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISN---DTVSPRA 115
```

```
Db         58 ----IPGETVHVRVPFASIQAVKVFLESQGIDYSIMIEDVQVLLDQEREEMLFNQQRERG 113

Qy        116 SASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAI 174

Db        114 GNFNFEAYHTLEEIYQEMDNLVAENPGLVSKVNLGSSFENRPMNVLKFSTGGD---KPAI 170

Qy        175 WIDCGIHAREWISPAFCLW-----------------------FI-------GH------ 197

Db        171 WLDAGIHAREWVTQATALWTANKIASDYGTDPAITSLLNTLDIFLLPVTNPDGYVFSQTT 230

Qy        198 NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 257

Db        231 NRMWRKTRSKRSGSGCVGVDPNRNW-DANFGGPGASSSPCSDSYHGPKPNSEVEVKSIVD 289

Qy        258 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYT 317

Db        290 FIKSH-GKVKAFITLHSYSQLLMFPYGYKCTKPDDFNELDEVAQKAAQAL-KRLHGTSYK 347

Qy        318 HGHGSETLYLAPGGGDDWIYDLGIKYSFTSN----------PPVEKLLP 356

Db        348 VGPICSVIYQASGGSIDWAYDLGIKYSFAFELRDTAFYGFLLPAKQILP 396
```

RESULT 9
CBPC_RAT
ID CBPC_RAT STANDARD; PRT; 309 AA.
AC P21961;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE A3).
GN CPA3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE.
RX MEDLINE=91105153; PubMed=1988052;
RA Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA Neurath H.;
RT "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT enzyme activity within mast cell granules.";
RL Biochemistry 30:648-655(1991).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC amino acid.
CC -!- SUBCELLULAR LOCATION: Secretory granules.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR PIR; A38395; A38395.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.010; -.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT METAL 68 68 ZINC (BY SIMILARITY).
FT METAL 71 71 ZINC (BY SIMILARITY).
FT METAL 196 196 ZINC (BY SIMILARITY).
FT ACT_SITE 248 248 PROTON DONOR (BY SIMILARITY).
FT ACT_SITE 270 270 NUCLEOPHILE (BY SIMILARITY).
FT DISULFID 65 78 BY SIMILARITY.
FT DISULFID 137 160 BY SIMILARITY.
FT VARIANT 1 1 MISSING (IN THE MAJOR FORM).
SQ SEQUENCE 309 AA; 35786 MW; 20330FABC3EE83EF CRC64;

Query Match 28.6%; Score 546; DB 1; Length 309;
Best Local Similarity 42.2%; Pred. No. 6.3e-39;
Matches 111; Conservative 39; Mismatches 73; Indels 40; Gaps 5;

```
Qy        120 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 179

Db          8 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 66

Qy        180 IHAREWISPAFCLWFI-------GHN-----------------------------RMWR 202

Db         67 IHAREWVSPAFCQWFVYQAAKSYGKNNIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 126

Qy        203 KNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262

Db        127 KNRSKNPSSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 185

Qy        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322

Db        186 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy        323 ETLYLAPGGGDDWIYDLGIKYSF 345

Db        245 STIYKTSGSSLDWAYDLGIKHTF 267
```

RESULT 10
CBPA_BOVIN
ID CBPA_BOVIN STANDARD; PRT; 419 AA.
AC P00730;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Carboxypeptidase A precursor (EC 3.4.17.1).
GN CPA.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91151335; PubMed=1998496;
RA le Hueerou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
RT "Cloning and nucleotide sequence of a bovine pancreatic
RT preprocarboxypeptidase A cDNA.";
RL Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96096549; PubMed=8522204;
RA Goo J.H., Kim K.H., Choi K.Y.;
RT "Cloning, sequencing and expression of the gene encoding a major
RT allotypic preprocarboxypeptidase A from bovine pancreas.";
RL Gene 165:333-334(1995).
RN [3]
RP SEQUENCE OF 111-417.
RX MEDLINE=71155183; PubMed=5102489;
RA Bradshaw R.A., Walsh K.A., Neurath H.;
RT "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT chymotryptic peptides of the cyanogen bromide fragment F-I.";
RL Biochemistry 10:938-950(1971).
RN [4]
RP REVISIONS TO 138 AND 141.
RX MEDLINE=72138789; PubMed=5143102;
RA Petra P.H., Hermodson M.A., Walsh K.A., Neurath H.;
RT "Characterization of bovine carboxypeptidase A (Allan).";
RL Biochemistry 10:4023-4025(1971).
RN [5]
RP SEQUENCE OF 17-120.
RX MEDLINE=89150306; PubMed=3147705;
RA Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neurath H.;
RT "The amino acid sequence of the activation peptide of bovine pro-
RT carboxypeptidase A.";
RL Biochimie 70:1137-1142(1988).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).

```
RX   MEDLINE=83294519; PubMed=6887246;
RA   Rees D.C., Lewis M., Lipscomb W.N.;
RT   "Refined crystal structure of carboxypeptidase A at 1.54-A
RT   resolution.";
RL   J. Mol. Biol. 168:367-387(1983).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX   MEDLINE=96003618; PubMed=7556081;
RA   Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT   "The three-dimensional structure of the native ternary complex of
RT   bovine pancreatic procarboxypeptidase A with proproteinase E and
RT   chymotrypsinogen C.";
RL   EMBO J. 14:4387-4394(1995).
RN   [8]
RP   VARIANT ALLELIC.
RX   MEDLINE=69283620; PubMed=5817619;
RA   Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Identification of the amino acid replacements characterizing the
RT   allotypic forms of bovine carboxypeptidase A.";
RL   Biochemistry 8:2762-2768(1969).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC       COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC       PROPROTEINASE E.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61851; AAA30426.1; -.
DR   EMBL; M61851; AAA30427.1; -.
DR   EMBL; Z33906; CAA83955.1; -.
DR   PIR; JN0126; CPBOA.
DR   PDB; 3CPA; 15-JAN-87.
DR   PDB; 4CPA; 22-OCT-84.
DR   PDB; 5CPA; 15-JAN-87.
DR   PDB; 6CPA; 15-OCT-91.
DR   PDB; 7CPA; 31-JAN-94.
DR   PDB; 8CPA; 31-JAN-94.
DR   PDB; 1CBX; 31-JAN-94.
DR   PDB; 1CPS; 15-OCT-94.
DR   PDB; 2CTB; 31-JAN-94.
DR   PDB; 2CTC; 31-JAN-94.
DR   PDB; 1ARL; 01-AUG-96.
DR   PDB; 1ARM; 17-AUG-96.
DR   PDB; 1BAV; 01-APR-97.
DR   PDB; 1YME; 12-FEB-97.
DR   PDB; 1CPX; 05-AUG-98.
DR   PDB; 1PYT; 27-JAN-97.
DR   PDB; 1ELL; 26-JUN-02.
DR   PDB; 1ELM; 26-JUN-02.
DR   PDB; 1F57; 06-SEP-00.
DR   PDB; 1HDQ; 15-NOV-01.
DR   PDB; 1HDU; 15-NOV-01.
DR   PDB; 1HEE; 23-NOV-01.
DR   PDB; 1M4L; 04-MAR-03.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure; Polymorphism.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A.
FT   METAL      179   179       ZINC.
FT   METAL      182   182       ZINC.
FT   METAL      306   306       ZINC.
FT   ACT_SITE   358   358       PROTON DONOR.
FT   ACT_SITE   380   380       NUCLEOPHILE.
FT   DISULFID   248   271
FT   VARIANT    289   289       I -> V (IN ALLELIC VARIANT).
FT   VARIANT    338   338       E -> A (IN ALLELIC VARIANT).
FT   VARIANT    415   415       L -> V (IN ALLELIC VARIANT).
FT   CONFLICT    95    95       S -> L (IN REF. 5).
FT   CONFLICT   199   199       D -> N (IN REF. 3).
FT   CONFLICT   203   203       D -> N (IN REF. 3).
FT   CONFLICT   224   224       D -> N (IN REF. 3).
FT   CONFLICT   232   232       Q -> E (IN REF. 3).
FT   CONFLICT   295   295       D -> N (IN REF. 3).
FT   TURN       114   116
FT   TURN       119   120
FT   HELIX      125   138
FT   TURN       140   142
FT   STRAND     143   150
FT   TURN       152   153
FT   STRAND     156   162
FT   STRAND     171   176
FT   TURN       180   181
FT   HELIX      183   199
FT   TURN       200   202
FT   HELIX      204   212
FT   STRAND     214   218
FT   HELIX      223   231
FT   TURN       232   232
FT   TURN       234   235
FT   STRAND     242   242
FT   TURN       244   245
FT   STRAND     249   249
FT   TURN       253   254
FT   TURN       261   262
FT   STRAND     267   267
FT   TURN       270   271
FT   TURN       273   274
FT   STRAND     275   275
FT   TURN       280   281
FT   HELIX      284   296
FT   STRAND     299   307
FT   STRAND     311   314
FT   TURN       324   325
FT   HELIX      326   344
FT   STRAND     349   352
FT   HELIX      353   356
FT   TURN       357   357
FT   HELIX      364   370
FT   TURN       371   372
FT   STRAND     375   381
FT   TURN       388   389
FT   HELIX      393   395
FT   HELIX      396   416
SQ   SEQUENCE   419 AA;  47082 MW;  21B86407B3BFC452 CRC64;

  Query Match             27.9%;  Score 533.5;  DB 1;  Length 419;
  Best Local Similarity   34.3%;  Pred. No. 1.1e-37;
  Matches  134;  Conservative   61;  Mismatches  133;  Indels   63;  Gaps   11;

Qy          3 LCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--W----QP 56
              |  |:||:   |  |     |   |||         :|| :: |     : |  |    ||
Db          4 LLILSVLLGAALGKED----FVGHQVLRITAADEAEVQTVKELEDLEHLQLDFWRGPGQP 59

Qy         57 VTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSP 113
                 :   |    :  ||  |   ||   ::: ||: |:   |:|:
Db         60 --------GSPIDVRVPFPSLQAVKVFLEAHGIRYRIMIEDVQSLLDEEQEQMFASQSRA 111
```

```
Qy        114 RASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKN 172
              |:: :: |  ||:|:|::| ::: :   || :::|: || |:|  |:|||| |    : :
Db        112 RSTNTFNYATYHTLDEIYDFMDLLVAEHPQLVSKLQIGRSYEGRPIYVLKFS-TGGSNRP 170

Qy        173 AIWIDCGIHAREWISPAFCLWFI---------------------------------- 195
              ||||| |||:||||: |  :||
Db        171 AIWIDLGIHSREWITQATGVWFAKKFTEDYGQDPSFTAILDSMDIFLEIVTNPDGFAFTH 230

Qy        196 GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAV 255
              ||:||| ||  ::: |:| | | |:    : : ||||| ||||| | |  || |||::
Db        231 SQNRLWRKTRSVTSSSLCVGVDANRNW-DAGFGKAGASSSPCSETYHGKYANSEVEVKSI 289

Qy        256 ASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTR 315
              |:: : |  ||::|:||||| :::|| ||     |  ||: ||  || |: |:   |
Db        290 VDFVKDHGN-FKAFLSIHSYSQLLLYPYGYTTQSIPDKTELNQVAKSAVEAL-KSLYGTS 347

Qy        316 YTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              | :|    |:| | ||  || |: |||||||
Db        348 YKYGSIITTIYQASGGSIDWSYNQGIKYSFT 378


RESULT 11
CBP1_RAT
ID   CBP1_RAT       STANDARD;      PRT;   419 AA.
AC   P00731;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82105986; PubMed=6275388;
RA   Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA   Pictet R.L., Valenzuela P., Rutter W.J.;
RT   "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT   characterization of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89034324; PubMed=3182872;
RA   Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT   "Structural characterization of the rat carboxypeptidase A1 and B
RT   genes. Comparative analysis of the rat carboxypeptidase gene
RT   family.";
RL   J. Biol. Chem. 263:17837-17845(1988).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01232; CAA24542.1; -.
DR   EMBL; J00713; AAA40893.1; -.
DR   EMBL; M23990; AAA40955.1; -.
DR   EMBL; M23960; AAA40955.1; JOINED.
DR   EMBL; M23985; AAA40955.1; JOINED.
DR   EMBL; M23986; AAA40955.1; JOINED.
DR   EMBL; M23987; AAA40955.1; JOINED.
DR   EMBL; M23988; AAA40955.1; JOINED.
DR   EMBL; M23989; AAA40955.1; JOINED.
DR   PIR; A00911; CPRTA.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16
FT   PROPEP      17    110       ACTIVATION PEPTIDE.
FT   CHAIN      111    419       CARBOXYPEPTIDASE A1.
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      182    182       ZINC (BY SIMILARITY).
FT   METAL      306    306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358    358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380    380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248    271       BY SIMILARITY.
FT   CONFLICT   196    196       I -> V (IN REF. 1).
FT   CONFLICT   261    263       FGM -> LGK (IN REF. 1).
FT   CONFLICT   347    347       K -> E (IN REF. 1).
SQ   SEQUENCE   419 AA;  47197 MW;  BB002D1CB99B7491 CRC64;

  Query Match             27.7%;  Score 529.5;  DB 1;  Length 419;
  Best Local Similarity   34.5%;  Pred. No. 2.4e-37;
  Matches  134;  Conservative   63;  Mismatches  136;  Indels   55;  Gaps   10;

Qy          2 KLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTA 59
              :|  |::|:  |   |     |  |||        ||| :: |     : |  |:
Db          3 RLLILSLLLEAVCGNEN----FVGHQVLRISAADEAQVQKVKELEDLEHLQLDFWRDAAR 58

Qy         60 DLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRAS 116
                |   :| |      : :||| |   ||   ::: ||: |:   :||:|       ::
Db         59 AGIPIDVRVPF----PSIQSVKAFLEYHGISYEIMIEDVQLLLDEEKQQMSAFQARALST 114

Qy        117 ASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIW 175
              |: |  ||:|:||| ::: :   || :::|| ||::||  |::||| |    | : |||
Db        115 DSFNYATYHTLDEIYEFMDLLVAEHPQLVSKIQIGNTFEGRPIHVLKFS-TGGTNRPAIW 173

Qy        176 IDCGIHAREWISPAFCLWFIGH-----------------------------------N 198
              || |||:|||:: |  :||                                       |
Db        174 IDTGIHSREWVTQASGVWFAKKITKDYGQDPTFTAVLDNMDIFLEIVTNPDGFAYTHKTN 233

Qy        199 RMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASF 258
              ||||| ||    : |:| | | |:    :   ||||: ||||| | :| || |||::  |
Db        234 RMWRKTRSHTQGSLCVGVDPNRNW-DAGFGMAGASSNPCSETYRGKFPNSEVEVKSIVDF 292

Qy        259 LRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTH 318
              :  : | |||:||:|||| :::|| ||   : |  ||  :|  || |:  :   |:: :
Db        293 VTSHGN-IKAFISIHSYSQLLLYPYGYTSEPAPDQAELDQLAKSAVTAL-TSLHGTKFKY 350

Qy        319 GHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |   :|:| | |   || |  |||||||
Db        351 GSIIDTIYQASGSTIDWTYSQGIKYSFT 378


RESULT 12
CPB2_HUMAN
ID   CPB2_HUMAN     STANDARD;      PRT;   417 AA.
AC   P48052; Q96A12; Q96QN3;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
```

```
RC   TISSUE=Pancreas;
RX   MEDLINE=95204457; PubMed=7896805;
RA   Catasus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
RA   Billeter M.;
RT   "The sequence and conformation of human pancreatic
RT   procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT   three-dimensional model.";
RL   J. Biol. Chem. 270:6651-6657(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98111000; PubMed=9450539;
RA   Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,
RA   Aviles F.X.;
RT   "Characterisation and preliminary X-ray diffraction analysis of human
RT   pancreatic procarboxypeptidase A2.";
RL   FEBS Lett. 420:7-10(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98046021; PubMed=9384570;
RA   Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT   "The three-dimensional structure of human procarboxypeptidase A2.
RT   Deciphering the basis of the inhibition, activation and intrinsic
RT   activity of the zymogen.";
RL   EMBO J. 16:6906-6913(1997).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19977; AAA74425.1; -.
DR   EMBL; BC007009; AAH07009.1; -.
DR   EMBL; BC014571; AAH14571.1; -.
DR   EMBL; BC015140; AAH15140.1; -.
DR   PIR; A56171; A56171.
DR   PDB; 1AYE; 13-JAN-99.
DR   PDB; 1DTD; 12-JUL-00.
DR   PDB; 1O6X; 30-JAN-03.
DR   MEROPS; M14.002; -.
DR   Genew; HGNC:2297; CPA2.
DR   MIM; 600688; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0007039; P:vacuolar protein catabolism; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   112       ACTIVATION PEPTIDE.
FT   CHAIN      113   417       CARBOXYPEPTIDASE A2.
FT   METAL      177   177       ZINC.
FT   METAL      180   180       ZINC.
FT   METAL      304   304       ZINC.
FT   ACT_SITE   378   378       NUCLEOPHILE.
FT   DISULFID   246   269
FT   DISULFID   318   352
FT   CONFLICT    80    80       E -> G (IN REF. 2; AAH07009).
FT   CONFLICT   302   302       T -> I (IN REF. 1).
SQ   SEQUENCE   417 AA;  46828 MW;  3D9C073A98700A83 CRC64;

  Query Match             26.7%;  Score 511;  DB 1;  Length 417;
  Best Local Similarity   32.5%;  Pred. No. 8.9e-36;
  Matches  128;  Conservative   64;  Mismatches  136;  Indels   66;  Gaps   11;

Qy         19 HVF---AFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTADLIVKKKQVHFFVN 73
              |::    |   |||  :|    |:: |  |     : |  |:  |       :  |  |
Db         13 HIYCLETFVGDQVLEIVPSNEEQIKNLLQLEAQEHLQLDFWKSPT----TPGETAHVRVP 68

Qy         74 ASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS---YYEQYHSLNEIY 130
                :|  ||  |   ||  |::: ||: |: ::      : |   |    :  ||:| ||
Db         69 FVNVQAVKVFLESQGIAYSIMIEDVQVLLDKENEEMLFNRRRERSGNFNFGAYHTLEEIS 128

Qy        131 SWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIHAREWISPA 189
               :: :   || :::|::|||||  |: ||| | | :   | |||:| |||||||:: |
Db        129 QEMDNLVAEHPGLVSKVNIGSSFENRPMNVLKFSTGGD---KPAIWLDAGIHAREWVTQA 185

Qy        190 FCLWFIG-----------------------------------HNRMWRKNRSFYANNH 212
               ||                                      |||||| ||  : :
Db        186 TALWTANKIVSDYGKDPSITSILDALDIFLLPVTNPDGYVFSQTKNRMWRKTRSKVSGSL 245

Qy        213 CIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISM 272
              |:| | | |:    :   ||||: ||::| |    || |||::  |:: :  ::||:|::
Db        246 CVGVDPNRNW-DAGFGGPGASSNPCSDSYHGPSANSEVEVKSIVDFIKSH-GKVKAFITL 303

Qy        273 HSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGG 332
              ||||| ::||| |  :|  | :||| || :| ::: ::   |:|  |     :| | ||
Db        304 HSYSQLLMFPYGYKCTKLDDFDELSEVAQKAAQSL-RSLHGTKYKVGPICSVIYQASGGS 362

Qy        333 DDWIYDLGIKYSFTSN----------PPVEKLLP 356
               || || ||||||              |  ::||
Db        363 IDWSYDYGIKYSFAFELRDTGRYGFLLPARQILP 396


RESULT 13
CBP1_HUMAN
ID   CBP1_HUMAN     STANDARD;      PRT;   419 AA.
AC   P15085; Q9BS67;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1 OR CPA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
```

```
RX   MEDLINE=93038569; PubMed=1417781;
RA   Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA   Wicker-Planquart C., Puigserver A.;
RT   "cDNA cloning and sequence analysis of human pancreatic
RT   procarboxypeptidase A1.";
RL   Biochem. J. 287:299-303(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 17-42.
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67318; CAA47732.1; -.
DR   EMBL; BC005279; AAH05279.1; -.
DR   PIR; S29127; S29127.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   Genew; HGNC:2296; CPA1.
DR   MIM; 114850; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16
FT   PROPEP      17    110       ACTIVATION PEPTIDE.
FT   CHAIN      111    419       CARBOXYPEPTIDASE A1.
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      182    182       ZINC (BY SIMILARITY).
FT   METAL      306    306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358    358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380    380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248    271       BY SIMILARITY.
FT   CONFLICT   276    276       H -> R (IN REF. 2).
SQ   SEQUENCE   419 AA;  47140 MW;  439FAFFFAEE958B1 CRC64;

  Query Match             25.6%;  Score 489.5;  DB 1;  Length 419;
  Best Local Similarity   32.7%;  Pred. No. 5.9e-34;
  Matches  131;  Conservative   60;  Mismatches  127;  Indels   83;  Gaps   13;

Qy          3 LCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQ----- 55
              |  |:||:  |   |      |   |||         ||| :: |     : |  |:
Db          4 LLVLSVLLGAVFGKED----FVGHQVLRISVADEAQVQKVKELEDLEHLQLDFWRGPAHP 59

Qy         56 --PVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQI---- 106
                |:          :| |       :  ||  |   ||    :: ||: |:   |:|:
Db         60 GSPIDV-------RVPF----PSIQAVKIFLESHGISYETMIEDVQSLLDEEQEQMFAFR 108

Qy        107 ----SNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLK 162
                  | || :       |  ||:| ||| ::: :   :| :::|| ||:::|  |:||||
Db        109 SRARSTDTFN-------YATYHTLEEIYDFLDLLVAENPHLVSKIQIGNTYEGRPIYVLK 161

Qy        163 VSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI-------------------------- 195
                |    : : ||||| |||:|||:: |  :||
Db        162 FS-TGGSKRPAIWIDTGIHSREWVTQASGVWFAKKITQDYGQDAAFTAILDTLDIFLEIV 220

Qy        196 ----------GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLY 245
                         |||||| ||  | : ||| | | |:    :   ||||: ||||| | :
Db        221 TNPDGFAFTHSTNRMWRKTRSHTAGSLCIGVDPNRNW-DAGFGLSGASSNPCSETYHGKF 279

Qy        246 PESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVR 305
                || |||::  |:: : | |||:||:||||| :::|| |     | :||  ::  ||
Db        280 ANSEVEVKSIVDFVKDHGN-IKAFISIHSYSQLLMYPYGYKTEPVPDQDELDQLSKAAVT 338

Qy        306 AIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
                |:  :   |:: :|   : :| | |   || |  |||||||
Db        339 AL-ASLYGTKFNYGSIIKAIYQASGSTIDWTYSQGIKYSFT 378


RESULT 14
CBP4_HUMAN
ID   CBP4_HUMAN     STANDARD;      PRT;   421 AA.
AC   Q9UI42;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN   CPA4 OR CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99310537; PubMed=10383164;
RA   Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT   "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT   deacetylase inhibitors during differentiation of prostate epithelial
RT   cancer cells.";
RL   Cancer Res. 59:2981-2988(1999).
RN   [2]
RP   IMPRINTING.
RX   MEDLINE=22483629; PubMed=12552318;
RA   Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA   Yoshiura K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA   Niikawa N., Kishino T.;
RT   "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT   imprinting domain.";
RL   Hum. Genet. 112:220-226(2003).
CC   -!- FUNCTION: Could be involved in the histone hyperacetylation
CC       pathway.
CC   -!- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,
```

```
CC       heart, intestine, kidney, liver and lung. Except for fetal brain
CC       that shows no imprinting, expression was found preferentially from
CC       the maternal allele.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095719; AAF23230.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.017; -.
DR   Genew; HGNC:15740; CPA4.
DR   MIM; 607635; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; NAS.
DR   GO; GO:0016573; P:histone acetylation; NAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   113       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN      114   421       CARBOXYPEPTIDASE A4.
FT   METAL      181   181       ZINC (BY SIMILARITY).
FT   METAL      184   184       ZINC (BY SIMILARITY).
FT   METAL      308   308       ZINC (BY SIMILARITY).
FT   ACT_SITE   382   382       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   250   273       BY SIMILARITY.
SQ   SEQUENCE   421 AA;  47379 MW;  91883D25B705BC6A CRC64;

  Query Match             25.3%;  Score 483;  DB 1;  Length 421;
  Best Local Similarity   32.0%;  Pred. No. 2.1e-33;
  Matches  119;  Conservative   63;  Mismatches  130;  Indels   60;  Gaps   10;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTA-----DLIVKKKQVHFFVNAS 75
              |   |||     |   ::  |  |  :  :  |  |:  ::     |::|    :  |
Db         21 FFGDQVLRINVRNGDEISKLSQLVNSNNLKLNFWKSPSSFNRPVDVLVPSVSLQAF---- 76

Qy         76 DVDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYS 131
                 |: |   |:  :| : |:: |:     :: ::    |:| :: |  ||||  ||
Db         77 -----KSFLRSQGLEYAVTIEDLQALLDNEDDEMQHNEGQERSSNNFNYGAYHSLEAIYH 131

Qy        132 WIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFC 191
               :: |    ||:  :: || |||  |:|||| |  :   : |:|:: |||:|||| |
Db        132 EMDNIAADFPDLARRVKIGHSFENRPMYVLKFSTGKGVRRPAVWLNAGIHSREWISQATA 191

Qy        192 LW-----------------------FI-----------GHNRMWRKNRSFYANNHCI 214
              :|                       |:             ||:||| ||    : ||
Db        192 IWTARKIVSDYQRDPAITSILEKMDIFLLPVANPDGYVYTQTQNRLWRKTRSRNPGSSCI 251

Qy        215 GTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHS 274
              | | | |: :  :  :||| : ||| | | :  || |||:|  |:::: |  | :| :||
Db        252 GADPNRNW-NASFAGKGASDNPCSEVYHGPHANSEVEVKSVVDFIQKHGN-FKGFIDLHS 309

Qy        275 YSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDD 334
              ||| :::|| |:  |: | |||  ||  | :|:   |  | |  |    |:| | |   |
Db        310 YSQLLMYPYGYSVKKAPDAEELDKVARLAAKALASVS-GTEYQVGPTCTTVYPASGSSID 368

Qy        335 WIYDLGIKYSFT 346
              | || |||::||
Db        369 WAYDNGIKFAFT 380
```

```
RESULT 15
CBPZ_SIMVI
ID   CBPZ_SIMVI     STANDARD;      PRT;   304 AA.
AC   P42788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS   Simulium vittatum (Black fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC   Simuliidae; Simulium.
OX   NCBI_TaxID=7192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Gut;
RX   MEDLINE=94093864; PubMed=8269093;
RA   Ramos A., Mahowald A., Jacobs-Lorena M.;
RT   "Gut-specific genes from the black fly Simulium vittatum encoding
RT   trypsin-like and carboxypeptidase-like proteins.";
RL   Insect Mol. Biol. 1:149-163(1993).
CC   -!- FUNCTION: Involved in the digestion of the blood meal.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Gut-specific.
CC   -!- INDUCTION: By blood meal.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08481; AAA18531.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.UPA; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   NON_TER      1     1
FT   METAL       58    58       ZINC (BY SIMILARITY).
FT   METAL       61    61       ZINC (BY SIMILARITY).
FT   METAL      184   184       ZINC (BY SIMILARITY).
FT   ACT_SITE   236   236       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   259   259       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   125   148       BY SIMILARITY.
SQ   SEQUENCE   304 AA;  34849 MW;  2E6E3FF8A6AA9144 CRC64;

  Query Match             21.2%;  Score 404.5;  DB 1;  Length 304;
  Best Local Similarity   33.0%;  Pred. No. 6e-27;
  Matches   93;  Conservative   52;  Mismatches   86;  Indels   51;  Gaps    9;

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              |||:| |||||:: : : ||: :  :  | |:|   :  :|||  |:      : ::  ||
Db          1 QYHTLPEIYSWLDRLVQEHPEHVEPVVGGKSYEGREIRGVKVSYKK--GNPVVMVESNIH 58

Qy        182 AREWISPA---FCL-------------------WFI-------GH------NRMWRKNRS 206
              |||||: |   : |                   |:|       |:      :|||||
Db         59 AREWITAATTTYLLNELLTSKNSTIREMAENYDWYIFPVTNPDGYVYTHTTDRMWRKTRS 118

Qy        207 FYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQI 266
                ::  | ||| |  |: :  || |:|  ||   |:||| |     || | :: : ||:    ||
Db        119 PNPDSLCAGTDPNRNW-NFHWMEQGTSSRPCTETYGGKKAFSEVETRSFSDFLKTLKGQI 177

Qy        267 KAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLY 326
              | |:: ||||| ::||| :|   : :|::|  :   | |:: :    | || |:  : :|
Db        178 KVYLAFHSYSQLLLFPYGHTCQHTYNHDDLQAIGDAAARSLAQ-RYGTDYTVGNIYDAIY 236
```

```
Qy        327 LAPGGGDDWIY---DLGIKYSFTSNP---------PVEKLLP 356
                  | ||  || |   |: | |::   |         |  :::|
Db        237 PASGGSMDWAYDTLDIPIAYTYELRPRDGWNGFQLPANQIIP 278
```

Search completed: January 1, 2004, 19:28:58
Job time : 16.2798 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:19:56 ; Search time 60.8768 Seconds
(without alignments)
1526.015 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1884	98.6	360	4	Q9P2Y6	Q9p2y6 homo sapien
2	1799.5	94.2	423	4	Q15114	Q15114 homo sapien
3	1794.5	93.9	423	4	Q96IY4	Q96iy4 homo sapien
4	1528	80.0	422	11	Q9JHH6	Q9jhh6 mus musculu
5	1522	79.6	422	11	Q9QZF0	Q9qzf0 mus musculu
6	1499	78.4	422	11	Q9EQV9	Q9eqv9 rattus norv
7	1029	53.8	198	4	Q9NTI8	Q9nti8 homo sapien
8	700.5	36.7	416	13	Q9PUF2	Q9puf2 bothrops ja
9	682	35.7	408	13	Q8AXN3	Q8axn3 paralichthy
10	646	33.8	416	6	Q9XSP3	Q9xsp3 sus scrofa
11	621	32.5	412	11	P97597	P97597 rattus norv
12	606	31.7	434	13	Q8AVS2	Q8avs2 xenopus lae
13	602	31.5	437	4	Q8N4T0	Q8n4t0 homo sapien
14	581	30.4	419	13	Q8UUK1	Q8uuk1 gallus gall
15	563	29.5	374	4	Q8IVL8	Q8ivl8 homo sapien
16	554.5	29.0	436	11	Q8R4H4	Q8r4h4 mus musculu
17	529.5	27.7	419	13	Q8AXN4	Q8axn4 paralichthy
18	519.5	27.2	436	4	Q8WXQ8	Q8wxq8 homo sapien
19	516.5	27.0	419	6	Q9TV85	Q9tv85 sus scrofa
20	506.5	26.5	420	11	Q8BMK6	Q8bmk6 mus musculu
21	484.5	25.4	452	13	Q8UW45	Q8uw45 fugu rubrip
22	477.5	25.0	262	11	Q8BVD0	Q8bvd0 mus musculu
23	471.5	24.7	247	4	Q96KZ9	Q96kz9 homo sapien
24	460.5	24.1	419	13	Q8AXN5	Q8axn5 paralichthy
25	460	24.1	424	5	Q9VL87	Q9vl87 drosophila
26	460	24.1	428	5	Q8IGC4	Q8igc4 drosophila
27	459	24.0	1132	5	Q9W475	Q9w475 drosophila
28	451.5	23.6	422	5	Q9VLZ2	Q9vlz2 drosophila
29	440.5	23.1	226	4	Q8TDE8	Q8tde8 homo sapien
30	436	22.8	584	5	Q9TZH1	Q9tzh1 caenorhabdi
31	434.5	22.7	351	5	Q8T0Q1	Q8t0q1 drosophila
32	434	22.7	351	4	Q8NA08	Q8na08 homo sapien
33	418	21.9	445	5	Q9W478	Q9w478 drosophila
34	416	21.8	488	5	Q9BL88	Q9bl88 caenorhabdi
35	412.5	21.6	416	5	Q9VRZ3	Q9vrz3 drosophila
36	407.5	21.3	419	5	Q9VLZ1	Q9vlz1 drosophila
37	402	21.0	424	5	Q9VS66	Q9vs66 drosophila
38	401.5	21.0	430	5	Q8SZP6	Q8szp6 drosophila
39	400.5	21.0	430	5	Q9VL86	Q9vl86 drosophila
40	399.5	20.9	427	5	Q9U9K2	Q9u9k2 aedes aegyp
41	397	20.8	323	5	Q19121	Q19121 caenorhabdi
42	395.5	20.7	540	5	Q9TZH2	Q9tzh2 caenorhabdi
43	391	20.5	453	5	Q9V342	Q9v342 drosophila
44	385	20.1	423	5	O61532	O61532 drosophila
45	382.5	20.0	581	5	Q23318	Q23318 caenorhabdi

ALIGNMENTS

RESULT 1
Q9P2Y6
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; AB011969; BAA90475.1; -.
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Query Match 98.6%; Score 1884; DB 4; Length 360;
Best Local Similarity 98.3%; Pred. No. 4.9e-162;
Matches 354; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQTGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSET 240
              |||||||||||||||||||||||||||||||||| || ||||||||||||||||||||||||
Db        181 HAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSET 240

Qy        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 YCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVA 300

Qy        301 SEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360
              |||||||:| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 SEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360
```

RESULT 2
Q15114
ID Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.009; -.
DR Genew; HGNC:2300; CPB2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;

Query Match 94.2%; Score 1799.5; DB 4; Length 423;
Best Local Similarity 89.6%; Pred. No. 2.7e-154;
Matches 343; Conservative 1; Mismatches 2; Indels 37; Gaps 1;

```
Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH----------------------------------NRMWRK 203
              |||||||||||||||||                                  ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383
```

RESULT 3
Q96IY4
ID Q96IY4 PRELIMINARY; PRT; 423 AA.
AC Q96IY4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B2 (Plasma).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC007057; AAH07057.1; -.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;

Query Match 93.9%; Score 1794.5; DB 4; Length 423;
Best Local Similarity 89.3%; Pred. No. 7.7e-154;
Matches 342; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

```
Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH----------------------------------NRMWRK 203
              |||||||||||||||||                                  ||||||
Db        181 HAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRK 240

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||||
Db        241 NRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 300

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              |||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||
```

```
Db        301 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSE 360

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              |||||||||||||||||||||||
Db        361 TLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 4
Q9JHH6
ID   Q9JHH6      PRELIMINARY;      PRT;   422 AA.
AC   Q9JHH6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE   (1110032P04Rik protein).
GN   CPB2 OR TAFI OR 1110032P04RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20341711; PubMed=10878383;
RA   Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA   Campbell W., Okada H.;
RT   "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT   whereas carboxypeptidase N is not.";
RL   J. Immunol. 165:1053-1058(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20201996; PubMed=10739389;
RA   Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA   van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT   "Characterization of mouse thrombin-activatable fibrinolysis
RT   inhibitor.";
RL   Thromb. Haemost. 83:297-303(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AB021968; BAB03402.1; -.
DR   EMBL; AF164524; AAF62385.1; -.
DR   EMBL; AK004045; BAB23141.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48870 MW;  99113755669D55CB CRC64;

  Query Match            80.0%;  Score 1528;  DB 11;  Length 422;
  Best Local Similarity  76.2%;  Pred. No. 9.9e-130;
  Matches  292;  Conservative   24;  Mismatches   29;  Indels   38;  Gaps    2;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              |||  | :|| |:|: ||| |||||||||:||||||||||:|||||||||:||||||||:
Db          1 MKLHGLGILVAIILY-EQHGFAFQSGQVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAE 59

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
                | |||:||||||||||:|||||||| || :||: :|||||:||  ||||||||||||
Db         60 FIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIEQQTFNDTVSPRASASYY 119

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||| |||:||||| ||:|||||||||||||||||||  ||||||||||
Db        120 EQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAIWIDCGI 179

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||:                                   ||||||
Db        180 HAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRK 239

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||| : || |:||||| || |||||:|||||||||||||||||||||||| ||||||
Db        240 NRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSETYCGLYPESEPEVKAVADFLRRNI 299

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              : |||||||||||| |:||||| |||||||||||||||||||:  :|||||||| |||
Db        300 DHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSE 359

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              :|||||| ||||||||||||||
Db        360 SLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 5
Q9QZF0
ID   Q9QZF0      PRELIMINARY;      PRT;   422 AA.
AC   Q9QZF0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase U.
GN   CPB2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/CBAF1J;
RA   He Y.C., Broze G.;
RT   "Isolation and characterization of mouse liver carboxypeptidase B
RT   gene.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF186188; AAF00528.1; -.
DR   HSSP; P00730; 5CPA.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48832 MW;  B53FFFB09943954E CRC64;

  Query Match            79.6%;  Score 1522;  DB 11;  Length 422;
  Best Local Similarity  76.0%;  Pred. No. 3.5e-129;
  Matches  291;  Conservative   24;  Mismatches   30;  Indels   38;  Gaps    2;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              |||  | :|| |:|: ||| |||||||||:||||||||||:|||||||||:||||||||:
```

```
Db          1 MKLHGLGILVAIILY-EQHGFAFQSGQVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAE 59

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
                | |||:||||||||||:|||||||| || :||: :|||||:||  ||||||||||||
Db         60 FIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIEQQTFNDTVSPRASASYY 119

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||| :||:||||| ||:||||||||||||||||||  ||||||||||
Db        120 EQYHSLNEIYSWIEVLTEQHPDMLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAIWIDCGI 179

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||:                                   ||||||
Db        180 HAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRK 239

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||| : || |:||||| || |||||:|||||||||||||||||||||||| ||||||
Db        240 NRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSETYCGLYPESEPEVKAVADFLRRNI 299

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              : |||||||||||| |:||||| |||||||||||||||||   :|||||||| |||
Db        300 DHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIGSFNKNTRYTHGSGSE 359

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              :|||||| ||||||||||||||
Db        360 SLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 6
Q9EQV9
ID   Q9EQV9      PRELIMINARY;      PRT;   422 AA.
AC   Q9EQV9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Pre-procarboxypeptidase R.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20471387; PubMed=11021404;
RA   Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA   Okada N., Okada H.;
RT   "Molecular cloning and partial characterization of rat
RT   procarboxypeptidase R and carboxypeptidase N.";
RL   Microbiol. Immunol. 44:719-728(2000).
DR   EMBL; AB042598; BAB18617.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   422 AA;  48826 MW;  FFFD32A51A9366C8 CRC64;

  Query Match             78.4%;  Score 1499;  DB 11;  Length 422;
  Best Local Similarity   74.7%;  Pred. No. 4.2e-127;
  Matches  286;  Conservative   26;  Mismatches   33;  Indels   38;  Gaps    2;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              |||  | ||| |:|: |:|  ||||| ||:||||||||||:||||||||:||||||||:
Db          1 MKLYGLGVLVAIILY-EKHGLAFQSGHVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAE 59

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
                | |||:||||||||::|||:|| | || :||: :|||||||| ||||||||||:|||
Db         60 FIEKKKEVHFFVNASDVNSVKAYLNASRIPFNVLMNNVEDLIQQQTSNDTVSPRASSSYY 119

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              ||||||||||||| |||:||||| ||:||||:|||||||||||||   ||||||||||
Db        120 EQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSYEKYPLYVLKVSGKEHRVKNAIWIDCGI 179

Qy        181 HAREWISPAFCLWFIGH-----------------------------------NRMWRK 203
              |||||||||||||||:                                   ||||||
Db        180 HAREWISPAFCLWFIGYVTQFHGKENTYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRK 239

Qy        204 NRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNI 263
              ||| : || |:||||| || |||||:|||| ||||||||||||||||||||| ||||||
Db        240 NRSVHMNNRCVGTDLNRNFASKHWCEKGASSFSCSETYCGLYPESEPEVKAVADFLRRNI 299

Qy        264 NQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSE 323
              | |||||||||||| |:||||| |||||||||||||||||:  :|||||||| |||
Db        300 NHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSE 359

Qy        324 TLYLAPGGGDDWIYDLGIKYSFT 346
              :|||||| ||||||||||||||
Db        360 SLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 7
Q9NTI8
ID   Q9NTI8      PRELIMINARY;      PRT;   198 AA.
AC   Q9NTI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN   CPB2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tromans A.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137141; CAB92622.1; -.
DR   HSSP; P00730; 1CPX.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
KW   Carboxypeptidase.
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  22460 MW;  32F005305621C2A5 CRC64;

  Query Match             53.8%;  Score 1029;  DB 4;  Length 198;
  Best Local Similarity   99.5%;  Pred. No. 4.9e-85;
  Matches  196;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTAD 60

Qy         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 LIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYY 120

Qy        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGI 180
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        121 EQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGI 180

Qy        181 HAREWISPAFCLWFIGH 197
              |||||||||||||||||
Db        181 HAREWISPAFCLWFIGH 197


RESULT 8
Q9PUF2
ID   Q9PUF2      PRELIMINARY;      PRT;   416 AA.
AC   Q9PUF2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
```

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase homolog.
OS   Bothrops jararaca (Jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
OX   NCBI_TaxID=8724;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT   "Screening of Bothrops jararaca pancreas cDNA library.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF190274; AAF01344.1; -.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ   SEQUENCE   416 AA;  47723 MW;  4F999854DD72B7A7 CRC64;

  Query Match             36.7%;  Score 700.5;  DB 13;  Length 416;
  Best Local Similarity   38.5%;  Pred. No. 7.3e-55;
  Matches  142;  Conservative   65;  Mismatches  119;  Indels   43;  Gaps    5;

Qy         15 FCEQHVFAFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNA 74
              | |  |  |   :|    ||   :|  |  |    ::  |:| : :|:  :  | | : |
Db         14 FAETTVHRFDGEKVYRVTPRNEDEVYFLNYLANIVQVDFWRPDSVELVKAEMTVDFRIEA 73

Qy         75 SDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIE 134
                |:: |  ||:   :|: ::: :: :|: |     | :   ||:|:|  :| :|
Db         74 DRCSEVESILQQSGLNYEILIDNLQAVLDRQLDNHA---RTAGYNYEKYNSWEKIDAWTA 130

Qy        135 FITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWF 194
               |   :| ::::: ||::||  |: :||| ||    | ||:|||| ||||||||||| ||
Db        131 DIANENPSLVSRLQIGTTFEGRPMPLLKV-GKPGVNKKAIFIDCGFHAREWISPAFCQWF 189

Qy        195 I-----------------------------------GHNRMWRKNRSFYANNHCIGTD 217
              :                                  :|||| ||  | : |||||
Db        190 VREAVRTYGKETIMTQLLNKLDFYILPVLNIDGYVYSWKQSRMWRKTRSVNAGSTCIGTD 249

Qy        218 LNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQ 277
               | ||    ||  ||| : |||||||  |||| | ||:| |:||| : |:||:::|||||
Db        250 PNRNF-DAAWCSVGASRNPCSETYCGSKPESEKETKALADFIRRNRSIIQAYLTIHSYSQ 308

Qy        278 HIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIY 337
              :::|||||   : ::::|: :| ||:| : |    | ||:| |: |:| | || ||| |
Db        309 MLLYPYSYTYDLTSNNKKLNSIAKEAIREL-KVLFGTEYTYGPGAATIYPAAGGSDDWAY 367

Qy        338 DLGIKYSFT 346
              | ||||:||
Db        368 DQGIKYAFT 376


RESULT 9
Q8AXN3
ID   Q8AXN3      PRELIMINARY;      PRT;   408 AA.
AC   Q8AXN3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B (Fragment).
GN   CPB.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Srivastava A.S., Kurokawa T., Suzuki T.;
RT   "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT   A2 and B from the Japanese flounder, Paralichthys olivaceus.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB099302; BAC53789.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   408 AA;  45904 MW;  809CAACD9D2DA34B CRC64;

  Query Match             35.7%;  Score 682;  DB 13;  Length 408;
  Best Local Similarity   39.3%;  Pred. No. 3.4e-53;
  Matches  142;  Conservative   61;  Mismatches  116;  Indels   42;  Gaps    5;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
              |:  :|    |      | ::::|  : |:  |:| :: |:    |   | |  :| |
Db         12 FEGEKVFRLKPVFDEHVTLIRDLANSIEVDFWRPESSQLVTIDIDVDIRVPAIYLDMVYT 71

Qy         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
                | | :   ||: |::  :  :     : |   |  | :|::|::: |||  |:  :||
Db         72 LLQQSDMEHEVLIEDLQSAVDAEAD---LKPSPRAHSYTKYNNLDKVQSWIASISSSNPD 128

Qy        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
               :::|  ||:::|  |: |||:  |  : | ||::||||||||||||||| ||:
Db        129 LISKQVIGNTYEGRPMTVLKLGKKSSSTKPAIFMDCGIHAREWISPAFCQWFVKEALSTY 188

Qy        196 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNSNFVSK 225
                                      |:      |||||| ||  : : ||| | | ||
Db        189 GSDSQMTSLLNQMDVFVLPVFNIDGYDFTHKSNRMWRKTRSRKSGSSCIGADPNRNF-DA 247

Qy        226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
               ||  ||| : ||:|:||  |||| ||| || |:||| : ||||:::|||||  ::|||||
Db        248 GWCSMGASDNPCSDTFCGYTPESEIEVKNVADFIRRNRSIIKAYLTVHSYSQLLLFPYSY 307

Qy        286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
               |   : || ||  ||  |  |: ::   | || | |: |:| | || ||| ||||:|||:
Db        308 TYQLAADHSELMKVAEGASAAL-RSLYGTTYTSGPGATTIYPAAGGSDDWAYDLGVKYSY 366

Qy        346 T 346
              |
Db        367 T 367


RESULT 10
Q9XSP3
ID   Q9XSP3      PRELIMINARY;      PRT;   416 AA.
AC   Q9XSP3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Procarboxypeptidase B precursor (EC 3.4.17.2).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=99321926; PubMed=10391940;
RA   Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA   Hershberger C., Aviles F.;
RT   "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT   Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL   J. Biol. Chem. 274:19925-19933(1999).
DR   EMBL; AJ133775; CAB46991.1; -.
DR   HSSP; P09955; 1PBA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
```

```
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Hydrolase; Signal.
FT   SIGNAL        1     15       POTENTIAL.
FT   CHAIN        16    416       PROCARBOXYPEPTIDASE B.
SQ   SEQUENCE   416 AA;  47381 MW;  5FA7FCED7B45AE6A CRC64;

  Query Match             33.8%;  Score 646;  DB 6;  Length 416;
  Best Local Similarity   36.6%;  Pred. No. 6.3e-50;
  Matches  140;  Conservative   64;  Mismatches  134;  Indels   44;  Gaps    7;

Qy          3 LCSLAVLVPIVLFCEQHVFA-FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADL 61
              : :  :|| : |     |    |:  :|          : :|  | :| :|  |:| :
Db          1 MLAFLILVTVTLASAHHSGEHFEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSVTQ 60

Qy         62 IVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYE 121
              |      | | | | |:  |:  |  : :   ||: ::  ::: |  :     | :   ||
Db         61 IKPHSTVDFRVKAEDILAVEDFLEQNELQYEVLINNLRSVLEAQFDSRV---RTTGHSYE 117

Qy        122 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 181
              :|::    | :| : :|  :||:::: ||::|    :|:||| ||    | ||::||| |
Db        118 KYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCGFH 176

Qy        182 AREWISPAFCLWFI------------------------------GH------NRMWRKN 204
              |||||| ||| ||:                              |:      ||||||
Db        177 AREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWRKT 236

Qy        205 RSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 264
              ||  |   ||||| | ||    ||  |||:  | |||||   ||| | ||:| |:| |::
Db        237 RSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNNLS 295

Qy        265 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSET 324
               ||||:::|||||  |::||||     ::: ||: :|  ||: :  |   |:||:| |: |
Db        296 SIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGATT 354

Qy        325 LYLAPGGGDDWIYDLGIKYSFT 346
              :| | || ||| || |||||||
Db        355 IYPAAGGSDDWAYDQGIKYSFT 376


RESULT 11
P97597
ID   P97597      PRELIMINARY;      PRT;   412 AA.
AC   P97597;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (Fragment).
GN   R-CPA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague Dawley;
RX   MEDLINE=97149430; PubMed=8996238;
RA   Lutzelschwab C., Pejler G., Aveskogh M., Hellman L.;
RT   "Secretory granule proteases in rat mast cells. Cloning of 10
RT   different serine proteases and a carboxypeptidase A from various rat
RT   mast cell populations.";
RL   J. Exp. Med. 185:13-29(1997).
DR   EMBL; U67914; AAB48267.1; -.
DR   HSSP; P09955; 1NSA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     10       POTENTIAL.
FT   CHAIN        11    412       MAST CELL CARBOXYPEPTIDASE A.
SQ   SEQUENCE   412 AA;  47944 MW;  DAB59555FC49137D CRC64;

  Query Match             32.5%;  Score 621;  DB 11;  Length 412;
  Best Local Similarity   37.2%;  Pred. No. 1.1e-47;
  Matches  134;  Conservative   55;  Mismatches  129;  Indels   42;  Gaps    6;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
              |   :|     :  :|  :|:||| | |:  | |      |     | | |   :  :::
Db         16 FDREKVFRVKLQDEKQASILKNLTQTIELDFWYPDAIHDIAVNMTVDFRVTEKESQTIQS 75

Qy         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142
              |     :  :|: |::: | :|  |      |     | :|:  |:| || | : :|:|:
Db         76 TLEQHKMDYEILINDLQEEIDKQF--DVKEEIAGRHSYAKYNDWNKIVSWTEKMVEKHPE 133

Qy        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 195
              |:::| |||: |  ||||||: |::   : ||::|||||||||:||||| ||:
Db        134 MVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCGIHAREWVSPAFCQWFVYQAAKSY 192

Qy        196 GHN----------------------------RMWRKNRSFYANNHCIGTDLNSNFVSK 225
              | |                            |||||||   |: ||||||| ||
Db        193 GKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWRKNRSKNPNSTCIGTDLNRNF-DV 251

Qy        226 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 285
              |     : : |   | |  |||| | ||| :|:| ::| |||||: ||||| ::||| |
Db        252 SWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLFPYGY 311

Qy        286 TRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345
              |     :|::|  ||  |   :  :  ||| :|  : |:|   |   || ||||||::|
Db        312 TIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIASTIYKTSGSSLDWAYDLGIKHTF 370


RESULT 12
Q8AVS2
ID   Q8AVS2      PRELIMINARY;      PRT;   434 AA.
AC   Q8AVS2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041315; AAH41315.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   434 AA;  50581 MW;  B42BD77843910591 CRC64;

  Query Match             31.7%;  Score 606;  DB 13;  Length 434;
  Best Local Similarity   35.0%;  Pred. No. 2.8e-46;
  Matches  137;  Conservative   63;  Mismatches  129;  Indels   62;  Gaps   10;

Qy          8 VLVPIVLFCEQHVFAFQSGQ-----------VLAALPRTSRQVQVLQNLTTTYEIVLWQ 55
              ||: ::| |     |  :||            |:   |:  |:   |:::    :: |||
Db         11 VLLFLLLSC-----ALSAGQCHLYNNRYSWDKVIRVFPKNEREANELKDMYQQLQVDLWQ 65

Qy         56 PVTADLIVKK--KQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQ-QQISNDTVS 112
              | :   | |      ||   |:|:|  :  :|  : |    :|: :|: ::: ||
Db         66 PSSISHIGKDTVTDVHTSGNSSEV--LLTYLTKAKIQHKILVNNVQSMLEAQQAFRPRRK 123

Qy        113 PRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAK 171
               |: : | | :|| |:|| ||: :: : | |:::   || |:|   |||||:    :: |
Db        124 RRSLSRYNYNEYHPLHEIESWMFYMNKTHHDLVSLFTIGKSYEGRSLYVLKLGKDTKSYK 183

Qy        172 NAIWIDCGIHAREWISPAFCLWFI------------------------------GH--- 197
```

```
Db        184 KAIWIDCGMHAREWIGPAFCQWFVKEAINSYNTDPAMKKILNLLYIYVMPVFNVDGYHYS 243

Qy        198 ---NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKA 254

Db        244 WHSDRFWRKTRSKNTRYQCYGVDANRNW-KVHWSDEGASLNPCDNTYCGPYAESEPEVKA 302

Qy        255 VASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNT 314

Db        303 VAQFLYKQRKHVRAYMSFHAYAQMLLYPYSYQYGAIPNFGCVESAAHNAVLAI-RSAYGI 361

Qy        315 RYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345

Db        362 RYRHGPASSTLYLTSGSSMDWAYNNGIPYSY 392


RESULT 13
Q8N4T0
ID   Q8N4T0      PRELIMINARY;      PRT;   437 AA.
AC   Q8N4T0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033684; AAH33684.1; -.
DR   Genew; HGNC:17245; CPA6.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   437 AA;  51024 MW;  9970680D429A0A98 CRC64;

  Query Match             31.5%;  Score 602;  DB 4;  Length 437;
  Best Local Similarity   34.4%;  Pred. No. 6.4e-46;
  Matches  134;  Conservative   67;  Mismatches  135;  Indels   54;  Gaps   10;

Qy          1 MKLC--SLAVLVPIVLFCEQHVF--AFQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQP 56

Db         15 LPLCWLFLKILQP----GHSHLYNNRYAGDKVIRFIPKTEEEAYALKKISYQLKVDLWQP 70

Qy         57 VTADLIVK--KKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPR 114

Db         71 SSISYVSEGTVTDVHIPQNGSRA--LLAFLQEANIQYKVLIEDLQKTLEKGSSLHTQRNR 128

Qy        115 ASAS--YYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKN 172

Db        129 RSLSGYNYEVYHSLEEIQNWMHHLNKTHSGLIHMFSIGRSYEGRCLFILKL-GRRSRLKR 187

Qy        173 AIWIDCGIHAREWISPAFCLWFI---------------------------------- 195

Db        188 AVWIDCGIHAREWIGPAFCQWFVKEALLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSW 247

Qy        196 GHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAV 255

Db        248 TNDRFWRKTRSRNSRFRCRGVDANRNWKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAV 306

Qy        256 ASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTR 315

Db        307 ANFLRKHRKHIRAYLSFHAYAQMLLYPYSYKYATIPNFRCVESAAYKAVNAL-QSVYGVR 365

Qy        316 YTHGHGSETLYLAPGGGDDWIYDLGIKYSF 345

Db        366 YRYGPASTTLYVSSGGSSMDWAYKNGIPYAF 395


RESULT 14
Q8UUK1
ID   Q8UUK1      PRELIMINARY;      PRT;   419 AA.
AC   Q8UUK1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Preprocarboxypeptidase A precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=domesticus; TISSUE=Pancreas;
RA   Hasegawa S., Honda K., Hikami Y.;
RT   "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
RT   precursor.";
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X64539; CAA45837.1; -.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Signal; Carboxypeptidase.
FT   SIGNAL        1     17       POTENTIAL.
FT   CHAIN       113    419       CARBOXYPEPTIDASE A.
SQ   SEQUENCE   419 AA;  47006 MW;  299C334274BCBC9C CRC64;

  Query Match             30.4%;  Score 581;  DB 13;  Length 419;
  Best Local Similarity   35.3%;  Pred. No. 4.8e-44;
  Matches  133;  Conservative   65;  Mismatches  107;  Indels   72;  Gaps    9;

Qy         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDN--- 79

Db         21 FVGHQVLRIVPSSDAELQKVQELQE--------------LEHLQLDFWLSPRGLGNPVD 65

Qy         80 ----------VKAHLNVSGIPCSVLLADVEDLI----QQQISNDTVSPRASASY-YEQYH 124

Db         66 IRVPFPSLQPVKAHLEANGVPYSIMIEDVQALVDLEQMQMLRRRRFVPLSTSTFDYTSYH 125

Qy        125 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 184

Db        126 TLEEIYAFMDLLVAENPNLVSKLEIGRTTENRPIYVLKFS-KGGTNRPAIWIDTGIHSRE 184

Qy        185 WISPAFCLWFI----------------------------------GHNRMWRKNRSFYA 209

Db        185 WVTQASGIWFAKKIVQEQDEGLANILDQMDIFLEIVTNPDGFAFTHTSNRMWRKTRSKRS 244

Qy        210 NNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAY 269

Db        245 GSLCVGVDPNRNW-DAGFGGSGASSNPCSETYHGPYANSEPEVKAIVDFVKSHKN-IKAF 302

Qy        270 ISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAP 329

Db        303 ISIHSYSQLLLYPYGYTTTAVPDKEELHQVAKEAVAAL-SSLYGTNYKYGSIITTIYQAS 361

Qy        330 GGGDDWIYDLGIKYSFT 346

Db        362 GGTIDWTYNQGIKYSFT 378


RESULT 15
```

```
Q8IVL8
ID   Q8IVL8      PRELIMINARY;      PRT;   374 AA.
AC   Q8IVL8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zn-carboxypeptidase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Obaya A.J., Lopez-Otin C.;
RT   "A new Zn-carboxypeptidase highly expressed in ovary.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ422118; CAD19478.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   374 AA;  42529 MW;  404C373BB841AAD2 CRC64;

  Query Match             29.5%;  Score 563;  DB 4;  Length 374;
  Best Local Similarity   38.7%;  Pred. No. 1.7e-42;
  Matches  109;  Conservative   51;  Mismatches   82;  Indels   40;  Gaps    5;

Qy        103 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVL 161
              :|:| : :||| :  :| |  || : ||| |:  |:|:: :::|:  :| ::| :|:| |
Db         29 RQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL 88

Qy        162 KVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------------------------- 195
              |:|      |  ||:||||||||||:|||| ||:
Db         89 KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLP 148

Qy        196 -----GH------NRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGL 244
                  |:      :|:|||:|| : |  | ||||| || :  ||  ||| :   :|:||
Db        149 VLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGT 207

Qy        245 YPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAV 304
              | |||| ||||||:    : |  :::|||| | |: || ||::|| :| |:  |  :|
Db        208 GPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAA 267

Qy        305 RAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
              |: |    | |  |  :: || : |   ||  |:|| :|:|
Db        268 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 308


Search completed: January  1, 2004, 19:31:55
Job time : 62.8768 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:33:22 ; Search time 393.069 Seconds
(without alignments)
2472.334 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135914_26819/app_query.fasta_1.1628
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_577_@runat_31122003_135914_26819 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1911	100.0	1573	21	AAC81962	Human brain carbox
2	1849.5	96.8	1625	25	ABX13670	Human protease cDN
3	1832	95.9	1272	20	AAV74302	Human plasma carbo
4	1832	95.9	1728	24	ABN95899	Gene #2397 used to
5	1832	95.9	1749	17	AAT11671	Human plasma carbo
6	1832	95.9	1749	18	AAT62846	Human plasma carbo
7	1827	95.6	1749	14	AAQ41001	Human plasma carbo
8	889	46.5	1400	21	AAF18005	Lung cancer associ
9	640.5	33.5	1215	16	AAQ90600	Porcine Tyr-His-Me
10	639	33.4	1263	17	AAT42494	Human pancreatic c
11	639	33.4	1263	19	AAV41795	Human pancreatic c
12	631.5	33.0	1284	17	AAT42506	ProHCPB gene with
13	623	32.6	2154	18	AAV17331	PreproHCPB-linker-
14	621.5	32.5	1332	24	ABL65831	Lung cancer relate
15	618.5	32.4	2023	25	ABZ75116	Anti-human seminal
16	611.5	32.0	1302	25	ABV77254	Nucleotide sequenc
17	611.5	32.0	1311	22	AAD03837	Human carboxypepti
18	611.5	32.0	1993	24	AAD30579	Human protease, PR
19	604	31.6	927	17	AAT35760	Rat mature carboxy
20	591	30.9	416	25	ABX49430	Bovine EST associa
21	576	30.1	921	16	AAQ90601	Porcine carboxypep
22	564	29.5	2128	22	AAD03839	Human carboxypepti
23	563.5	29.5	1378	25	ABT33355	NOVX DNA sequence
24	563	29.5	1125	22	AAH76477	Human secreted met
25	563	29.5	1125	22	AAH76479	cDNA encoding huma
26	563	29.5	1125	24	ABK31745	DNA encoding novel
27	563	29.5	1222	25	ABT33353	NOVX DNA sequence
28	563	29.5	1342	22	AAH76476	Human secreted met
29	563	29.5	1342	22	AAH76478	Human secreted met
30	563	29.5	1344	25	ABQ77393	Human zinc-binding
31	563	29.5	1603	24	AAD39053	Human 23566 (carbo
32	563	29.5	1826	22	AAF81728	Human protease and
33	561	29.4	999	17	AAT42500	Mature HCPB coding
34	561	29.4	1053	17	AAT42497	mature HCPB-(His)6
35	554	29.0	1053	20	AAX24804	Human carboxypepta
36	554	29.0	1059	17	AAT42511	Modified HCPB (D25
37	554	29.0	1059	18	AAT62787	Carboxypeptidase B
38	553	28.9	1059	17	AAT42512	Modified HCPB (D25
39	553	28.9	1059	18	AAT62788	Carboxypeptidase B
40	550	28.8	1059	18	AAT62806	Carboxypeptidase B
41	548	28.7	1059	18	AAT62790	Carboxypeptidase B
42	548	28.7	1059	18	AAT62801	Carboxypeptidase B
43	547	28.6	1059	18	AAT62789	Carboxypeptidase B
44	547	28.6	1196	25	ABT33352	NOVX DNA sequence
45	546	28.6	1059	18	AAT62791	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.
XX
AC AAC81962;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B cDNA.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             18..1100
FT                   /*tag= a
FT                   /product= "carboxypeptidase B"
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
DR   P-PSDB; AAB11457.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Claim 2b; Page 64-68; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

Alignment Scores:
Pred. No.:               7.2e-195      Length:          1573
Score:                   1911.00       Matches:         360
Percent Similarity:      100.00%       Conservative:    0
Best Local Similarity:   100.00%       Mismatches:      0
Query Match:             100.00%       Indels:          0
DB:                      21            Gaps:            0

US-09-980-881A-2 (1-360) x AAC81962 (1-1573)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         18 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 77

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 TTCGCGTTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 137

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 197

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 257

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        258 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 317

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        318 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 377

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        378 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 437

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        438 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 497

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        498 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 557

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 617

Qy        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGC 677

Qy        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 AACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 737

Qy        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        738 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 797

Qy        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        798 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 857

Qy        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        858 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 917

Qy        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        918 AGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCAT 977

Qy        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        978 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1037

Qy        341 IleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1038 ATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097

RESULT 2
ABX13670
ID   ABX13670 standard; cDNA; 1625 BP.
XX
AC   ABX13670;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease cDNA.
XX
KW   Human; gene; ss; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..16
FT                   /*tag= a
FT   CDS             17..1177
FT                   /*tag= b
FT                   /product= "Protease"
FT   variation       replace(521,A)
```

```
FT                   /*tag=  c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR           1178..1625
FT                   /*tag=  d
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   P-PSDB; ABG72093, ABG72095.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 4; Fig 1; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   cDNA, the gene for which is located on chromosome 13.
XX
SQ   Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

Alignment Scores:
Pred. No.:              3.01e-188      Length:         1625
Score:                  1849.50        Matches:        355
Percent Similarity:     94.18%         Conservative:   1
Best Local Similarity:  93.92%         Mismatches:     4
Query Match:            96.78%         Indels:         18
DB:                     25             Gaps:           1

US-09-980-881A-2 (1-360) x ABX13670 (1-1625)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 316

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 556

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 616

Qy        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 676

Qy        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 736

Qy        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 796

Qy        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 856

Qy        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320
              ||||||||||||||||||||||:::|||  ||||||||||||||||||||||||||||||||
Db        917 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 976

Qy        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1036

Qy        341 IleLysTyrSerPhe---------------------------------------------- 345
              |||||||||||||||
Db       1037 ATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCG 1096

Qy        346 --------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                      ||||||||||||||||||||||||||||||||||||||||||||
Db       1097 GAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1148

RESULT 3
AAV74302
ID   AAV74302 standard; cDNA; 1272 BP.
```

```
XX
AC   AAV74302;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW   polymorphism detection; thrombotic disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   P-PSDB; AAW92270.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Example 1; Page 24; 35pp; English.
XX
CC   This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Alignment Scores:
Pred. No.:              1.59e-186      Length:          1272
Score:                  1832.00        Matches:         357
Percent Similarity:     86.27%         Conservative:    1
Best Local Similarity:  86.02%         Mismatches:      2
Query Match:            95.87%         Indels:          55
DB:                     20             Gaps:            2

US-09-980-881A-2 (1-360) x AAV74302 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        197 ------------------------------------------------------------ 197

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 4
ABN95899
```

```
ID   ABN95899 standard; DNA; 1728 BP.
XX
AC   ABN95899;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2397 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 2397; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;

Alignment Scores:
Pred. No.:                2.49e-186      Length:          1728
Score:                    1832.00        Matches:         357
Percent Similarity:       86.27%         Conservative:    1
Best Local Similarity:    86.02%         Mismatches:      2
Query Match:              95.87%         Indels:          55
DB:                       24             Gaps:            2

US-09-980-881A-2 (1-360) x ABN95899 (1-1728)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        197 ------------------------------------------------------------ 197

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        198 -----------------------------------------AsnArgMetTrpArgLys 203
                                                        ||||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1160 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1219

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1220 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262
```

```
RESULT 5
AAT11671
ID   AAT11671 standard; DNA; 1749 BP.
XX
AC   AAT11671;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human plasma carboxypeptidase B.
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   P-PSDB; AAR90293.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:
Pred. No.:              2.54e-186      Length:         1749
Score:                  1832.00        Matches:        357
Percent Similarity:     86.27%         Conservative:   1
Best Local Similarity:  86.02%         Mismatches:     2
Query Match:            95.87%         Indels:         55
DB:                     17             Gaps:           2

US-09-980-881A-2 (1-360) x AAT11671 (1-1749)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
                 ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        197 ------------------------------------------------------------ 197

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                        ||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
                 ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy        344 SerPhe----------------------------------------------------Th 346
                  ||||||                                                ||
Db       1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                  ||||||||||||||||||||||||||||||||||||||||
Db       1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 6
AAT62846
ID   AAT62846 standard; DNA; 1749 BP.
XX
AC   AAT62846;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human PCPB
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
FT   misc_binding    134..177
FT                   /*tag=  d
FT                   /bound_moeity= 46_bp_probe
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   P-PSDB; AAW14733.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence encodes human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:

Pred. No.:              2.54e-186      Length:          1749
Score:                  1832.00        Matches:         357
Percent Similarity:     86.27%         Conservative:    1
Best Local Similarity:  86.02%         Mismatches:      2
Query Match:            95.87%         Indels:          55
DB:                     18             Gaps:            2

US-09-980-881A-2 (1-360) x AAT62846 (1-1749)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        197 ------------------------------------------------------------ 197

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000
```

```
Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy        344 SerPhe----------------------------------------------------Th 346
              ||||||                                                    ||
Db       1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              ||||||||||||||||||||||||||||||||||||||||||||
Db       1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283
```

```
RESULT 7
AAQ41001
ID   AAQ41001 standard; cDNA; 1749 BP.
XX
AC   AAQ41001;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B gene.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     41..106
FT                   /*tag=  a
FT   mat_peptide     107..1312
FT                   /*tag=  b
FT                   /product= PCPB
FT   misc_feature    134..177
FT                   /*tag=  c
FT                   /function= probe
FT                   /note= "used to obtain full-length clones"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   P-PSDB; AAR36273.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Disclosure; Fig 4; 40pp; English.
XX
CC   Human plasma carboxypeptidase B was isolated from human plasma and
CC   partially sequenced. Oligonucleotide primers were designed based on
CC   the partial amino acid sequences. The primers were used in a PCR
CC   amplification to identify cDNA encoding PCPB from a human liver cDNA
CC   library. The PCR product was capable of encoding the first 37 amino
CC   acids of PCPB; a 46mer probe was used to obtain the full-length
CC   sequence which, although disclosed in the specification, is not
CC   claimed. PCPB inhibits the enzymatic conversion by tPA of
CC   plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;
```

```
Alignment Scores:
Pred. No.:                8.73e-186      Length:          1749
Score:                    1827.00        Matches:         356
Percent Similarity:       86.02%         Conservative:    1
Best Local Similarity:    85.78%         Mismatches:      3
Query Match:              95.60%         Indels:          55
DB:                       14             Gaps:            2

US-09-980-881A-2 (1-360) x AAQ41001 (1-1749)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy        197 ------------------------------------------------------------ 197

Db        641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy        198 -----------------------------------------AsnArgMetTrpArgLys 203
                                                        ||||||||||||||||
Db        701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        821 TCCAAACACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy        344 SerPhe-------------------------------------------------------Th 346
              ||||||                                                       ||
Db       1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1241 ATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAA 1283

RESULT 8
AAF18005
ID   AAF18005 standard; DNA; 1400 BP.
XX
AC   AAF18005;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58129.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 1; Page 507; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

Alignment Scores:
Pred. No.:              3e-85          Length:          1400
Score:                  889.00         Matches:         178
Percent Similarity:     75.21%         Conservative:    1
Best Local Similarity:  74.79%         Mismatches:      3
Query Match:            46.52%         Indels:          56
DB:                     21             Gaps:            2

US-09-980-881A-2 (1-360) x AAF18005 (1-1400)

Qy        179 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 197
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         10 GGAATCCATGCCAGANAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATA 69

Qy        197 ------------------------------------------------------------ 197

Db         70 ACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTAT 129

Qy        198 ----------------------------------------------AsnArgMetTr 201
                                                           ||||||||||
Db        130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACTACTCATGGAAAAAGAATCGAATGTG 189

Qy        201 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAs 221
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||    ||
Db        190 GAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAA 249

Qy        221 nPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 241
              ||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309

Qy        241 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 261
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAG 369

Qy        261 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 281
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 AAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTT 429

Qy        281 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 TCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAG 489

Qy        301 rGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 321
              ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db        490 TGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGG 549

Qy        321 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIl 341
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609

Qy        341 eLysTyrSerPhe------------------------------------------------ 345
              |||||||||||
Db        610 CAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGA 669

Qy        346 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                   ||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 719

RESULT 9
```

```
AAQ90600
ID   AAQ90600 standard; DNA; 1215 BP.
XX
AC   AAQ90600;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB; ds.
XX
OS   Sus scrofa.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1215
FT                   /*tag=  a
FT                   /product= Porcine procarboxypeptidase B.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   P-PSDB; AAR75131.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Claim 2; Page 20-21; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

Alignment Scores:
Pred. No.:              1.14e-58        Length:         1215
Score:                  640.50          Matches:        136
Percent Similarity:     54.29%          Conservative:   60
Best Local Similarity:  37.67%          Mismatches:     122
Query Match:            33.52%          Indels:         43
DB:                     16              Gaps:           6

US-09-980-881A-2 (1-360) x AAQ90600 (1-1215)

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              |||:::      :::|||                                :::   :::|||
Db         28 TTCGAAGGGGAGAAGGTGTTCCGTGTCAATGTTGAAGATGAAAATGACATCAGCTTACTC 87

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
                    |||   :::|||   :::|||      |||:::|||   :::          |||
Db         88 CATGAGTTGGCCAGCACCAGGCAGATTGACTTCTGGAAACCAGATTCTGTCACACAAATC 147

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                           |||   |||   |||   |||   |||:::      |||:::
Db        148 AAACCTCACAGTACAGTTGACTTCCGTGTGAAAGCAGAAGATATTTTGGCTGTGGAAGAC 207

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
              |||      :::  :::          |||||:::   ::::::      ::::::
Db        208 TTTCTGGAGCAGAATGAACTACAATATGAGGTACTCATAAACAACCTGAGATCTGTGCTC 267

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
              :::   |||      :::                |||   :::         ||||||:::
Db        268 GAGGCTCAGTTTGACAGCAGAGTC---------CGTACAACTGGACACAGTTATGAGAAG 318

Qy        123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
              |||::::::         |||   :::|||   :::   :::|||      :::||||||
Db        319 TACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAAAATCCAGAC 378

Qy        143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
              ::::::::::::      ||||||::::::|||           :::|||:::||||||
Db        379 CTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATACCTCCTCAAG 438

Qy        163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
              |||   ||||||             |||   ||||||::::::||||||||   ||||||
Db        439 GTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 495

Qy        183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 195
              ||||||||||||||   |||||||||   ||||||:::
Db        496 AGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTTCTCACCTAT 555

Qy        195 ------------------------------------------------------------ 195

Db        556 GGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTCTTGCCTGTG 615

Qy        196 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 205
                        |||:::                 ||||||||||||||||   |||
Db        616 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCGC 675

Qy        206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
              |||      |||          |||||||||||||   |||   ||||||
Db        676 TCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 732

Qy        226 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245
              ||||||      |||||||||:::        |||   ||||||||||||||
Db        733 GGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGCT 792

Qy        246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265
              |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        793 GCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTCC 852

Qy        266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 285
              |||||||||||:::::::::|||||||||||||||   |||::::::||||||||||||
Db        853 ATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTAT 912

Qy        286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 305
                       :::::::::   ||||||:::   :::|||      ||||||:::
Db        913 GATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAAA 972

Qy        306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325
                 :::      |||          |||:::||||||:::|||   |||:::   |||:::
Db        973 GAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAATC 1029

Qy        326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345
              |||   |||   ||||||   |||||||||   ||||||   ||||||||||||||||||
Db       1030 TATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTTC 1089

Qy        346 Thr 346
              |||
Db       1090 ACC 1092

RESULT 10
AAT42494
ID   AAT42494 standard; DNA; 1263 BP.
XX
AC   AAT42494;
```

```
XX
DT   12-FEB-1997  (first entry)
XX
DE   Human pancreatic carboxypeptidase B encoding sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1248
FT                   /*tag=  a
FT                   /product= preproenzyme
FT   sig_peptide     1..39
FT                   /*tag=  b
FT                   /note= "partial enzyme secretion leader sequence
FT                        (pre-sequence); residues -108 to -96"
FT   sig_peptide     40..324
FT                   /*tag=  c
FT                   /note= "pro-sequence; residues -95 to -1"
FT   mat_peptide     385..1245
FT                   /*tag=  d
FT                   /note= "mature enzyme; encodes residues +1 to +307"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06172.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 14; Page 131; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence encodes the native prepro-HCPB (see
CC   AAW06172) contained in vector pICI1698.
XX
SQ   Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;

Alignment Scores:
Pred. No.:              1.76e-58      Length:         1263
Score:                  639.00        Matches:        135
Percent Similarity:     53.83%        Conservative:   69
Best Local Similarity:  35.62%        Mismatches:     131
Query Match:            33.44%        Indels:         44
DB:                     17            Gaps:           7

US-09-980-881A-2 (1-360) x AAT42494 (1-1263)
```

```
Qy          6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla---PheGln 24
              |||   |||||||||   :::   |||         |||           |||:::
Db          4 CTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAGCACTTTGAA 63

Qy         25 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 44
                 :::|||                          :::   :::::::::
Db         64 GGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATCCGCGAG 123

Qy         45 LeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLys 64
              |||   :::|||   :::|||       |||:::|||   :::         |||
Db        124 TTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATCAAACCT 183

Qy         65 LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 84
                   |||   |||   |||   |||   |||         |||:::   |||
Db        184 CACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAATGTTCTA 243

Qy         85 AsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 104
                 :::   :::         ||||||:::::::::::::   ::::::::::::
Db        244 AAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTGGAGGCT 303

Qy        105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHis 124
              |||       :::               ||||||:::         ||||||:::|||:::
Db        304 CAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAGTACAAC 354

Qy        125 SerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeu 144
                       |||   :::|||   :::   :::         :::|||   ::::::
Db        355 AAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCCCTCATC 414

Qy        145 ThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
              ::::::      ||||||::::::||||||          :::|||:::|||||||||
Db        415 TCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAGGTT--- 471

Qy        165 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGlu 184
              ||||||             |||   ||||||::::::|||||||   ||||||||||
Db        472 GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCCAGAGAG 531

Qy        185 TrpIleSerProAlaPheCysLeuTrpPheIle-------------------------- 195
              ||||||||||||||||||||   ||||||:::
Db        532 TGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTATGGACGT 591

Qy        195 ---------------------------------------------------------- 195

Db        592 GAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTGCTCAAT 651

Qy        196 ------GlyHis-----------------AsnArgMetTrpArgLysAsnArgSerPhe 207
                    |||:::                 :::|||   |||||||||   ||||||
Db        652 ATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGCTCCACC 711

Qy        208 TyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
              :::       :::   ||||||||||||||||   |||   ||||||            |||
Db        712 CATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCTGGTTGG 768

Qy        228 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 247
              ||||||   |||||||||   :::   |||   ||||||||||||||||          |||
Db        769 TGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCCGCAGAG 828

Qy        248 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
              ||||||   |||   ||||||:::|||   |||:::|||       ::::::   ||||||
Db        829 TCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCCATCAAG 888

Qy        268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyrThrArg 287
              ||||||:::::::::|||||||||||||||   :::::::::||||||||||||
Db        889 GCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATATGCTTAC 948

Qy        288 SerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArgAlaIle 307
                    :::::::::   ||||||:::   :::|||         |||:::   :::
Db        949 AAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAAGAACTT 1008

Qy        308 AspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
                  :::         |||:::||||||:::|||   |||:::   |||:::|||
Db       1009 ---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATCTATCCT 1065

Qy        328 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346
```

```
Db       1066 GCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTCACC 1122

RESULT 11
AAV41795
ID   AAV41795 standard; DNA; 1263 BP.
XX
AC   AAV41795;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Human pancreatic carboxypeptidase nucleotide sequence.
XX
KW   ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW   prodrug therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1248
FT                   /*tag=  a
FT                   /product= "Pancreatic carboxypeptidase B"
FT                   /note= "No start codon given"
XX
PN   WO9835988-A1.
XX
PD   20-AUG-1998.
XX
PF   10-FEB-1998;   98WO-GB00415.
XX
PR   29-OCT-1997;   97GB-0022727.
PR   14-FEB-1997;   97GB-0003104.
PR   18-OCT-1997;   97GB-0022003.
XX
PA   (ZENE ) ZENEPA   (ZENE ) ZENECA LTD.
XX
PI   Edge MD;
XX
DR   WPI; 1998-467168/40.
DR   P-PSDB; AAW74476.
XX
PT   New modified pro-domain of carboxy-peptidase B - enhances expression
PT   of co-expressed proteins for production of recombinant
PT   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT   enzyme prodrug therapy
XX
PS   Example 1; Page 52-53; 83pp; English.
XX
CC   The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC   modified pro-domain of CPB on a separate gene to enhance recombinant
CC   expression.  This process can be used to produce recombinant CPB in
CC   eukaryotic cells, or fusions of CPB with antibody chains.  CPB is used in
CC   insulin production and protein sequencing, while its fusions with
CC   antibodies are useful in antibody-directed enzyme prodrug therapy.  The
CC   modified pro-domain provide increased yields of recombinant CPB, possibly
CC   by protecting the C-terminus against enzymatic degradation or by
CC   increasing intracellular trafficking.
XX
SQ   Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

Alignment Scores:
Pred. No.:              1.76e-58      Length:          1263
Score:                  639.00        Matches:         135
Percent Similarity:     53.83%        Conservative:    69
Best Local Similarity:  35.62%        Mismatches:      131
Query Match:            33.44%        Indels:          44
DB:                     19            Gaps:            7

US-09-980-881A-2 (1-360) x AAV41795 (1-1263)

Qy          6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla---PheGln 24
Db          4 CTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAGCACTTTGAA 63

Qy         25 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 44
Db         64 GGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATCCGCGAG 123

Qy         45 LeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLys 64
Db        124 TTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATCAAACCT 183

Qy         65 LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 84
Db        184 CACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAATGTTCTA 243

Qy         85 AsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 104
Db        244 AAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTGGAGGCT 303

Qy        105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHis 124
Db        304 CAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAGTACAAC 354

Qy        125 SerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeu 144
Db        355 AAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCCCTCATC 414

Qy        145 ThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
Db        415 TCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAGGTT--- 471

Qy        165 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGlu 184
Db        472 GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCCAGAGAG 531

Qy        185 TrpIleSerProAlaPheCysLeuTrpPheIle---------------------------- 195
Db        532 TGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTATGGACGT 591

Qy        195 ------------------------------------------------------------ 195
Db        592 GAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTGCTCAAT 651

Qy        196 ------GlyHis-----------------AsnArgMetTrpArgLysAsnArgSerPhe 207
Db        652 ATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGCTCCACC 711

Qy        208 TyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
Db        712 CATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCTGGTTGG 768

Qy        228 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 247
Db        769 TGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCCGCAGAG 828

Qy        248 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
Db        829 TCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCCATCAAG 888

Qy        268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyrThrArg 287
Db        889 GCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATATGCTTAC 948

Qy        288 SerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArgAlaIle 307
Db        949 AAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAAGAACTT 1008

Qy        308 AspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
Db       1009 ---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATCTATCCT 1065

Qy        328 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346
Db       1066 GCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTCACC 1122
```

```
RESULT 12
AAT42506
ID   AAT42506 standard; DNA; 1284 BP.
XX
AC   AAT42506;
XX
DT   17-FEB-1997  (first entry)
XX
DE   ProHCPB gene with PelB leader sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1275
FT                   /*tag=  a
FT   sig_peptide     1..66
FT                   /*tag=  b
FT                   /note= "PelB leader sequence"
FT   sig_peptide     67..351
FT                   /*tag=  c
FT                   /note= "pro sequence"
FT   mat_peptide     352..1272
FT                   /*tag=  d
FT                   /note= "mature HCPB sequence"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JP, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06175.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 18; Page 140-142; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is a cloned pro-HCPB gene contained in
CC   plasmid pICI1738 and which can be expressed in E. coli.
XX
SQ   Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;

Alignment Scores:
Pred. No.:                1.15e-57      Length:          1284
Score:                    631.50        Matches:         129
Percent Similarity:       54.57%        Conservative:    68
Best Local Similarity:    35.73%        Mismatches:      121
Query Match:              33.05%        Indels:          43
DB:                       17            Gaps:            6

US-09-980-881A-2 (1-360) x AAT42506 (1-1284)

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              |||:::     :::|||                          :::   ::::::
Db         85 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 144

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db        145 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 204

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                          |||   |||   |||   |||   |||          |||:::
Db        205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 264

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
              |||       :::   :::          ||||||:::::::::::::   :::::::::
Db        265 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
              :::   |||      :::                 ||||||:::         ||||||:::
Db        325 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 375

Qy        123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
              |||:::            |||   :::|||   :::   :::         :::|||
Db        376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 435

Qy        143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
              :::::::::::      ||||||:::::||||||        :::|||:::||||||
Db        436 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 495

Qy        163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
              |||   ||||||            |||   ||||||:::::|||||||||   ||||||
Db        496 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552

Qy        183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 195
              ||||||||||||||||||||||||||||||||   ||||||:::
Db        553 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 612

Qy        195 ------------------------------------------------------------ 195

Db        613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 672

Qy        196 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 205
                        |||:::                 :::|||   |||||||||   |||
Db        673 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 732

Qy        206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
              |||   :::      :::   |||||||||||||   |||   ||||||
Db        733 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy        226 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245
              |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db        790 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849

Qy        246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265
              |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        850 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 909

Qy        266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 285
              ||||||||||:::::::::|||||||||||||||   :::::::::||||||||||||
Db        910 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 969

Qy        286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 305
                          :::::::::   ||||||:::   :::|||         |||:::
Db        970 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy        306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325
                 :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db       1030 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1086

Qy        326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345
              |||   |||   ||||||   |||||||||   ||||||   ||||||:::|||||||||
```

```
Db       1087 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146

Qy        346 Thr 346
                |||
Db       1147 ACC 1149

RESULT 13
AAV17331
ID   AAV17331 standard; DNA; 2154 BP.
XX
AC   AAV17331;
XX
DT   02-JUN-1998  (first entry)
XX
DE   PreproHCPB-linker-humanised 806.077 VH gene.
XX
KW   Anti-CEA antibody; carcinoembryonic antigen; 806.077 Ab; cancer therapy;
KW   cancer diagnosis; complementarity determining region; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2151
FT                   /*tag=  a
XX
PN   WO9742329-A1.
XX
PD   13-NOV-1997.
XX
PF   29-APR-1997;   97WO-GB01165.
XX
PR   14-FEB-1997;   97GB-0003103.
PR   04-MAY-1996;   96GB-0009405.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Copley CG,  Edge MD,  Emery SC;
XX
DR   WPI; 1997-558987/51.
DR   P-PSDB; AAW41414.
XX
PT   Anti-carcinoembryonic antigen antibody 806.077 Ab - used for
PT   diagnosis and therapy of cancer
XX
PS   Example 75; Page 181-183; 208pp; English.
XX
CC   This sequence encodes preproHCPB-linker-humanised 807.077 VH, which is
CC   an example of the antibody of the invention. The antibody is an anti-CEA
CC   (carcinoembryonic antigen) antibody (preferably 806.077 Ab). Host cells
CC   or transgenic organisms transformed with DNA encoding the antibody, are
CC   used to make the antibody or conjugate. The conjugate is used in a
CC   medicament suitable for intravenous administration. The conjugate can be
CC   used for cancer therapy, selectively killing tumour cells. The antibody
CC   can be used for in vivo or in vitro diagnosis of cancer.
XX
SQ   Sequence 2154 BP; 532 A; 601 C; 548 G; 473 T; 0 other;

Alignment Scores:
Pred. No.:              2.01e-56        Length:         2154
Score:                  623.00          Matches:        132
Percent Similarity:     53.40%          Conservative:   72
Best Local Similarity:  34.55%          Mismatches:     134
Query Match:            32.60%          Indels:         44
DB:                     18              Gaps:           7

US-09-980-881A-2 (1-360) x AAV17331 (1-2154)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              :::   :::|||   |||||||||   :::   |||             |||
Db          1 ATGTTGGCACTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAG 60

Qy         23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
                   |||:::      :::|||                                :::   :::

Db         61 CACTTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATA 120

Qy         42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61
              ::::::   |||   :::|||   :::|||       |||:::|||   :::
Db        121 ATCCGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAA 180

Qy         62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
              |||                |||   |||   |||   |||   |||          |||:::
Db        181 ATCAAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAG 240

Qy         82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
                 |||       :::   :::         ||||||:::::::::::   ::::::
Db        241 AATGTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTG 300

Qy        102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ::::::   |||       :::                |||||:::         ||||||
Db        301 GTGGAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAG 351

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              :::|||:::             |||   :::|||   :::   :::          :::|||
Db        352 AAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCA 411

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
                 ::::::::::::       ||||||::::::||||||        :::|||:::|||
Db        412 GCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTG 471

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              ||||||   ||||||            |||   ||||||::::::|||||||||   |||
Db        472 AAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCAT 528

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 195
              ||||||||||||||||||||||||||||||||   |||||||::
Db        529 GCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACC 588

Qy        195 ------------------------------------------------------------ 195

Db        589 TATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCT 648

Qy        196 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204
                           |||:::                 :::|||   |||||||||
Db        649 GTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACT 708

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||   :::       :::   ||||||||||||||   |||   ||||||
Db        709 CGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GAT 765

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
                 |||||||||   |||||||||  :::   |||   |||||||||||||||
Db        766 GCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCT 825

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
                 |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db        826 GCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCT 885

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284
              |||||||||||:::::::::|||||||||||||||   :::::::::||||||||||
Db        886 TCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCA 945

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              |||                :::::::::   ||||||:::   :::|||          |||
Db        946 TATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTG 1005

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
                 :::   :::       :::         |||:::||||||:::|||   |||:::   |||
Db       1006 AAAGAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACA 1062

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              :::|||   :::   |||         ||||||   ||||||   ||||||:::||||||
Db       1063 ATCTATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTATGACCAAGGAATCAGATATTCC 1122

Qy        345 PheThr 346
              ||||||
Db       1123 TTCACC 1128
```

```
RESULT 14
ABL65831
ID   ABL65831 standard; DNA; 1332 BP.
XX
AC   ABL65831;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Lung cancer related gene sequence SEQ ID NO:4168.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set  -
XX
PS   Claim 1; SEQ ID 4168; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 1332 BP; 357 A; 321 C; 311 G; 343 T; 0 other;

Alignment Scores:
Pred. No.:              1.43e-56        Length:         1332
Score:                  621.50          Matches:        134
Percent Similarity:     53.66%          Conservative:   71
Best Local Similarity:  35.08%          Mismatches:     132
Query Match:            32.52%          Indels:         45
DB:                     24              Gaps:           8

US-09-980-881A-2 (1-360) x ABL65831 (1-1332)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              :::   :::|||   |||||||||   :::  |||           |||
Db         24 ATGTTGGCACTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAG 83

Qy         23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
                 |||:::     :::|||                          :::   :::
Db         84 CACTTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATA 143

Qy         42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61
              ::::::   |||   :::|||   :::|||      |||:::|||   :::
Db        144 ATCCGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAA 203

Qy         62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
              |||                |||   |||   |||   |||   |||         |||:::
Db        204 ATCAAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAG 263

Qy         82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
                 |||      :::   :::       ||||||:::::::::::::   ::::::
Db        264 AATGTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTG 323

Qy        102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ::::::   |||      :::                |||||:::          ||||||
Db        324 GTGGAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAG 374

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              :::|||:::             |||   :::|||   :::   :::         :::|||
Db        375 AAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCA 434

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
                 ::::::::::::      ||||||::::::||||||        :::|||:::|||
Db        435 GCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTG 494

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              ||||||   ||||||            |||   ||||||::::::||||||||   |||
Db        495 AAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCAT 551
```

```
Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 195
              ||||||||||||||||||||||||||||||||   ||||||:::
Db        552 GCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACC 611

Qy        195 ------------------------------------------------------------ 195

Db        612 TATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCT 671

Qy        196 ---------------GlyHis------------------AsnArgMetTrpArgLysAsn 204
                          |||:::                  :::|||   |||||||||
Db        672 GTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACT 731

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||   :::      :::      ||||||||||||   |||   ||||||
Db        732 CGCTCCACCCATACTGGATCT---AGCATTGGCACAGACCCCAACAGAAATTTT---GAT 785

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
                  ||||||||   |||||||||   :::   |||   ||||||||||||||
Db        786 GCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCT 845

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
                ||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        846 GCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCT 905

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284
              ||||||||||||:::::::::|||||||||||||||   :::::::::||||||||||
Db        906 TCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCA 965

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              |||                :::::::::   |||||:::   :::|||         |||
Db        966 TATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTG 1025

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
              :::   :::      :::        |||:::||||||:::|||   |||:::   |||
Db       1026 AAAGAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACA 1082

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              :::|||   |||   ||||||   |||||||||   ||||||   ||||||:::|||||||
Db       1083 ATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCC 1142

Qy        345 PheThr 346
              ||||||
Db       1143 TTCACC 1148
```

RESULT 15
ABZ75116
ID ABZ75116 standard; DNA; 2023 BP.
XX
AC ABZ75116;
XX
DT 30-MAY-2003 (first entry)
XX
DE Anti-human seminal plasma protein ScFv/hCPA fusion gene.
XX
KW Single chain antibody; ScFv; anti-human seminal plasma protein;
KW monoclonal antibody E4B7; mouse; murine; human; carboxypeptidase A;
KW hCPA; fusion protein; antibody-directed enzyme prodrug therapy; ADEPT;
KW methotrexate-alpha-peptide prodrug; prostate cancer; cytostatic;
KW drug targeting; heavy chain variable region; VH; VL;
KW light chain variable region; gene; ds.
XX
OS Chimeric - Mus musculus.
OS Chimeric - Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 7..2010
FT /*tag= a
FT /product= "Anti-human seminal plasma protein single-chain
FT antibody (ScFv)/human carboxypeptidase A
FT (hCPA) fusion protein"
FT /transl_except= (pos:652..654,aa:Asn)
FT /transl_except= (pos:745..747,aa:Cys)
FT mat_peptide 7..741
FT /*tag= b
FT /note= "Encodes ScFv portion of fusion protein"
FT misc_feature 742..759
FT /*tag= c
FT /note= "Linker DNA"
FT mat_peptide 760..2007
FT /*tag= d
FT /note= "Encodes hCPA portion of fusion protein"
XX
PN WO2002100431-A1.
XX
PD 19-DEC-2002.
XX
PF 08-JUN-2001; 2001WO-CN00924.
XX
PR 08-JUN-2001; 2001WO-CN00924.
XX
PA (UYTA-) UNIV TANGDU HOSPITAL FORTH MILITARY MEDI.
PA (HAOX/) HAO X.
XX
PI Hao X;
XX
DR WPI; 2003-156924/15.
DR P-PSDB; ABP97414.
XX
PT Pharmaceutical kits for use in the treatment of prostate cancer,
PT comprise anti-human seminal plasma protein single-chain antibody/human
PT carboxypeptidase fusion protein and a prodrug of
PT methotrexate-alpha-peptides -
XX
PS Claim 4; Fig 3; 29pp; Chinese.
XX
CC The invention relates to a pharmaceutical kit for antibody-directed
CC enzyme prodrug therapy (ADEPT) to treat prostate cancer. The kit
CC comprises several respective containers for holding anti-human seminal
CC plasma protein single-chain antibody (ScFv)/human carboxypeptidase A
CC (hCPA) fusion protein (ABP97414), a prodrug of methotrexate-alpha-peptide
CC and a pharmaceutically-acceptable auxiliary for medication. The invention
CC also encompasses the anti-human seminal plasma protein ScFv/hCPA fusion
CC protein, and the nucleic acid encoding it (ABZ75116). This was
CC constructed by linking DNA encoding the ScFv (itself generated by linking
CC DNAs encoding the heavy and light chain variable regions (VH and VL) of
CC the murine anti-human seminal plasma protein monoclonal antibody E4B7)
CC to DNA encoding human carboxypeptidase A via a linker. The ScFv portion
CC of the fusion protein enables it to be specifically targeted to prostate
CC cancer cells, where the hCPA portion can then activate the prodrug which
CC in turn has cytostatic activity towards the cancer cells. The invention
CC permits the fusion protein to be easily produced in large quantities for
CC use in ADEPT for prostate cancer treatment. The present sequence
CC represents DNA encoding the ScFv/hCPA fusion protein of the invention.
XX
SQ Sequence 2023 BP; 581 A; 465 C; 467 G; 510 T; 0 other;

Alignment Scores:

Pred. No.:	5.57e-56	Length:	2023
Score:	618.50	Matches:	130
Percent Similarity:	54.09%	Conservative:	75
Best Local Similarity:	34.30%	Mismatches:	127
Query Match:	32.37%	Indels:	47
DB:	25	Gaps:	8

US-09-980-881A-2 (1-360) x ABZ75116 (1-2023)

```
Qy          8 ValLeuValProIleValLeuPheCysGluGlnHisValPheAla---------PheGln 24
              ::::::::::|||:::   |||                  |||         |||
Db        763 CTCATCCTGCCTGTGGGTTTGATTGCTACCACTCTTGCAATTGCTCCTGTCCGCTTTGAC 822

Qy         25 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 44
                    :::|||           |||:::      :::|||      ::::::::::::
Db        823 AGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATAAAGGAC 882

Qy         45 LeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLys 64
```

```
Db        883 TTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTAGCTGCT 942

Qy         65 LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 84

Db        943 AATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCTGCCTTG 1002

Qy         85 AsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 104

Db       1003 GATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATTGAGAAA 1062

Qy        105 GlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyr 123

Db       1063 CAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCAAAATAC 1113

Qy        124 HisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAspMet 143

Db       1114 AATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCTGAAATG 1173

Qy        144 LeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysVal 163

Db       1174 GTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTGAAGATT 1233

Qy        164 SerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArg 183

Db       1234 ---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCACGCACGA 1290

Qy        184 GluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 195

Db       1291 GAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACTTATGGG 1350

Qy        195 ------------------------------------------------------------ 195

Db       1351 AGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCTGTGTTC 1410

Qy        196 ---------GlyHis------------------AsnArgMetTrpArgLysAsnArgSer 206

Db       1411 AATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAATCGTTCC 1470

Qy        207 PheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHis 226

Db       1471 AAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AATGCTTCA 1527

Qy        227 TrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrPro 246

Db       1528 TGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCTGCACCA 1587

Qy        247 GluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIle 266

Db       1588 GAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAATGAAATC 1647

Qy        267 LysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyrThr 286

Db       1648 AAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGATATACA 1707

Qy        287 ArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArgAla 306

Db       1708 TCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACTGATGTT 1767

Qy        307 IleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyr 326

Db       1768 CTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACAATTTAC 1824

Qy        327 LeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345

Db       1825 CCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACATTT 1881


Search completed: January 1, 2004, 20:09:01
Job time : 419.069 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:51:12 ; Search time 165.344 Seconds
(without alignments)
961.011 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135917_26894/app_query.fasta_1.1628
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_115_@runat_31122003_135917_26894 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1849.5	96.8	1625	4	US-09-813-133A-1	Sequence 1, Appli
	2	1832	95.9	1272	2	US-08-869-057-1	Sequence 1, Appli
	3	1832	95.9	1749	1	US-07-649-591B-2	Sequence 2, Appli
	4	1832	95.9	1749	1	US-08-277-540-2	Sequence 2, Appli
	5	1832	95.9	1749	1	US-08-430-787A-2	Sequence 2, Appli
	6	640.5	33.5	1215	1	US-08-696-139-1	Sequence 1, Appli
	7	639	33.4	1263	2	US-08-860-882A-56	Sequence 56, Appl
	8	639	33.4	1263	4	US-09-011-769A-38	Sequence 38, Appl
	9	631.5	33.0	1284	2	US-08-860-882A-71	Sequence 71, Appl
	10	631.5	33.0	1284	4	US-09-011-769A-55	Sequence 55, Appl
	11	623	32.6	2154	3	US-09-171-945-124	Sequence 124, App
	12	611.5	32.0	1311	4	US-09-675-305-9	Sequence 9, Appli
	13	604	31.6	927	2	US-08-782-760-5	Sequence 5, Appli
	14	604	31.6	927	5	PCT-US96-00995-5	Sequence 5, Appli
	15	576	30.1	921	1	US-08-696-139-3	Sequence 3, Appli
	16	564	29.5	2128	4	US-09-675-305-13	Sequence 13, Appl
	17	561	29.4	999	2	US-08-860-882A-67	Sequence 67, Appl
	18	561	29.4	999	4	US-09-011-769A-50	Sequence 50, Appl
	19	561	29.4	1053	2	US-08-860-882A-64	Sequence 64, Appl
	20	561	29.4	1053	4	US-09-011-769A-46	Sequence 46, Appl
	21	554	29.0	1053	4	US-09-463-451-27	Sequence 27, Appl
c	22	554	29.0	1053	4	US-09-463-451-28	Sequence 28, Appl
	23	554	29.0	1059	2	US-08-860-882A-74	Sequence 74, Appl
	24	554	29.0	1059	4	US-09-011-769A-59	Sequence 59, Appl
	25	553	28.9	1059	2	US-08-860-882A-77	Sequence 77, Appl
	26	553	28.9	1059	4	US-09-011-769A-63	Sequence 63, Appl
	27	546	28.6	1200	4	US-09-710-099-7	Sequence 7, Appli
	28	543	28.4	1870	3	US-09-171-945-112	Sequence 112, App
	29	528.5	27.7	1050	4	US-09-675-305-11	Sequence 11, Appl
	30	519.5	27.2	1311	4	US-09-710-099-5	Sequence 5, Appli
	31	513	26.8	1251	3	US-08-640-906-3	Sequence 3, Appli
	32	513	26.8	1251	4	US-09-395-936-3	Sequence 3, Appli
	33	487.5	25.5	1257	3	US-08-640-906-1	Sequence 1, Appli
	34	487.5	25.5	1257	4	US-09-395-936-1	Sequence 1, Appli
	35	460.5	24.1	945	4	US-09-710-099-3	Sequence 3, Appli
	36	460.5	24.1	945	4	US-09-710-099-11	Sequence 11, Appl
	37	458.5	24.0	2247	4	US-09-710-099-15	Sequence 15, Appl
	38	434	22.7	1056	4	US-09-710-099-1	Sequence 1, Appli
	39	434	22.7	1056	4	US-09-710-099-9	Sequence 9, Appli
	40	363	19.0	55827	4	US-09-813-133A-3	Sequence 3, Appli
	41	289.5	15.1	741	4	US-09-675-305-5	Sequence 5, Appli
	42	213	11.1	591	3	US-09-331-709-2	Sequence 2, Appli
	43	198.5	10.4	629	4	US-09-280-116-228	Sequence 228, App
c	44	168.5	8.8	515	3	US-08-998-416-125	Sequence 125, App
	45	162.5	8.5	494	4	US-09-669-751-63	Sequence 63, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Alignment Scores:
Pred. No.: 2.02e-221 Length: 1625
Score: 1849.50 Matches: 355
Percent Similarity: 94.18% Conservative: 1
Best Local Similarity: 93.92% Mismatches: 4
Query Match: 96.78% Indels: 18
DB: 4 Gaps: 1

US-09-980-881A-2 (1-360) x US-09-813-133A-1 (1-1625)

```
Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         17 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 76

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db         77 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 136

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        137 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 196

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        197 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 256

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        257 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 316

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        317 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 376

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        377 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 436

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        437 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 496

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              |||||||||||||||||||||||  |||||||||||||||||||||||||||||||||||
Db        497 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 556

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMet 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        557 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATG 616

Qy        201 TrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSer 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        617 TGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGG 676

Qy        221 AsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThr 240
              ||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        677 AACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACC 736

Qy        241 TyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArg 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        737 TACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGA 796

Qy        261 ArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        797 AGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTG 856

Qy        281 PheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAla 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        857 TTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCC 916

Qy        301 SerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHis 320
              ||||||||||||||||||:::|||  ||||||||||||||||||||||||||||||||||||
Db        917 AGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCAT 976

Qy        321 GlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGly 340
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        977 GGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGC 1036

Qy        341 IleLysTyrSerPhe---------------------------------------------- 345
              |||||||||||||||
Db       1037 ATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCG 1096

Qy        346 -------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                    ||||||||||||||||||||||||||||||||||||||||||||||
Db       1097 GAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1148

RESULT 2
US-08-869-057-1
; Sequence 1, Application US/08869057
```

```
; Patent No. 5985562
; GENERAL INFORMATION:
;   APPLICANT: Morser, Michael J
;   APPLICANT: Nagashima, Mariko
;   TITLE OF INVENTION: Method of Detecting Thrombotic Disease
;   TITLE OF INVENTION: Risk
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Berlex Biosciences Legal Department
;     STREET: 15049 San Pablo Avenue
;     CITY: Richmond
;     STATE: California
;     COUNTRY: USA
;     ZIP: 94804-0099
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/869,057
;     FILING DATE: 03-JUN-1997
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Washtien, Wendy L
;     REGISTRATION NUMBER: 36,301
;     REFERENCE/DOCKET NUMBER: 51509AUSM1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 510-262-5411
;     TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1272 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   PUBLICATION INFORMATION:
;     AUTHORS: Eaton, Dan L.
;     AUTHORS: Malloy, Beth E.
;     AUTHORS: Tsai, Siao P
;     AUTHORS: Henzel, William
;     AUTHORS: Drayna, Dennis
;     TITLE: Isolation, Molecular Cloning, and Partial
;     TITLE: Characterization of a No. 5985562el Carboxypeptidase B
;     TITLE: from Human Plasma
;     JOURNAL: J. Biol. Chem.
;     VOLUME: 266
;     ISSUE: 32
;     PAGES: 21833-21838
;     DATE: No. 5985562 15-1991
US-08-869-057-1

Alignment Scores:
Pred. No.:              2.08e-219      Length:          1272
Score:                  1832.00        Matches:         357
Percent Similarity:     86.27%         Conservative:    1
Best Local Similarity:  86.02%         Mismatches:      2
Query Match:            95.87%         Indels:          55
DB:                     2              Gaps:            2

US-09-980-881A-2 (1-360) x US-08-869-057-1 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal  80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 240

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 300

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 420

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 480

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 540

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 600

Qy        197 ------------------------------------------------------------ 197

Db        601 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 660

Qy        198 -----------------------------------------AsnArgMetTrpArgLys 203
                                                        ||||||||||||||||||
Db        661 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 780

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 840

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 900

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 960

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1080

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy        344 SerPhe-----------------------------------------------------Th 346
              ||||||                                                     ||
Db       1141 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1200

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1201 ATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAA 1243
```

```
RESULT 3
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  hybridization probe
;      LOCATION:  133 to 178
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  potential clip site
;      LOCATION:  380 to 382
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  signal sequence
;      LOCATION:  41 to 106
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-07-649-591B-2

Alignment Scores:
Pred. No.:                3.58e-219       Length:          1749
Score:                    1832.00         Matches:         357
Percent Similarity:       86.27%          Conservative:    1
Best Local Similarity:    86.02%          Mismatches:      2
Query Match:              95.87%          Indels:          55
DB:                       1               Gaps:            2

US-09-980-881A-2 (1-360) x US-07-649-591B-2 (1-1749)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100
```

```
Qy        21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160

Qy        41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy       197 ------------------------------------------------------------ 197

Db       641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy       198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db       701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy       204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy       224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy       244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy       264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy       284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy       304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
             ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db      1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy       324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       344 SerPhe---------------------------------------------------Th 346
```

```
             ||||||                                                    ||
Db      1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy       346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
             |||||||||||||||||||||||||||||||||||||||||||||
Db      1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 4
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-277-540-2

Alignment Scores:
Pred. No.:              3.58e-219      Length:          1749
Score:                  1832.00        Matches:         357
Percent Similarity:     86.27%         Conservative:    1
Best Local Similarity:  86.02%         Mismatches:      2
Query Match:            95.87%         Indels:          55
DB:                     1              Gaps:            2

US-09-980-881A-2 (1-360) x US-08-277-540-2 (1-1749)

Qy         1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy        21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160
```

```
Qy        41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy       197 ------------------------------------------------------------ 197

Db       641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy       198 -----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||||
Db       701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy       204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy       224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy       244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy       264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy       284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy       304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
             ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy       324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       344 SerPhe-------------------------------------------------------Th 346
             ||||||                                                       ||
Db      1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240
```

```
Qy       346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
             ||||||||||||||||||||||||||||||||||||||||||||
Db      1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-787A-2

Alignment Scores:
Pred. No.:              3.58e-219       Length:         1749
Score:                  1832.00         Matches:        357
Percent Similarity:     86.27%          Conservative:   1
Best Local Similarity:  86.02%          Mismatches:     2
Query Match:            95.87%          Indels:         55
DB:                     1               Gaps:           2

US-09-980-881A-2 (1-360) x US-08-430-787A-2 (1-1749)

Qy         1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        41 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 100

Qy        21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       101 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 160
```

```
Qy        41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       161 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 220

Qy        61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       221 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 280

Qy        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       281 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 340

Qy       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       341 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 400

Qy       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       401 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 460

Qy       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       461 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 520

Qy       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       521 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 580

Qy       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db       581 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 640

Qy       197 ------------------------------------------------------------ 197

Db       641 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 700

Qy       198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db       701 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 760

Qy       204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       761 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 820

Qy       224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       821 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 880

Qy       244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       881 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 940

Qy       264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       941 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 1000

Qy       284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1001 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1060

Qy       304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
             ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||||
Db      1061 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1120

Qy       324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1121 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1180

Qy       344 SerPhe-------------------------------------------------Th 346
             ||||||                                                 ||
Db      1181 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1240

Qy       346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
```

```
             ||||||||||||||||||||||||||||||||||||||||
Db      1241 ATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 6
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
;   APPLICANT:  Fayerman, Jeffrey T.
;   APPLICANT:  Greenen, David P.
;   APPLICANT:  Hershberger, Charles L.
;   APPLICANT:  Larson, Jeffrey L.
;   APPLICANT:  Sterner, Jane L.
;   APPLICANT:  Zhang, Haichao
;   TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;   TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;   NUMBER OF SEQUENCES:  6
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Eli Lilly and Company
;     STREET:  Lilly Corporate Center
;     CITY:  Indianapolis
;     STATE:  Indiana
;     COUNTRY:  United States of America
;     ZIP:  46285
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/696,139
;     FILING DATE:
;     CLASSIFICATION:  435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/153,258
;     FILING DATE:  16-NOV-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Gaylo, Paul J.
;     REGISTRATION NUMBER:  36,808
;     REFERENCE/DOCKET NUMBER:  X-8681
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (317) 276-0756
;     TELEFAX:  (317) 276-3861
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  1215 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   FEATURE:
;     NAME/KEY:  CDS
;     LOCATION:  1..1215
US-08-696-139-1

Alignment Scores:
Pred. No.:              3.85e-70      Length:          1215
Score:                  640.50        Matches:         136
Percent Similarity:     54.29%        Conservative:    60
Best Local Similarity:  37.67%        Mismatches:      122
Query Match:            33.52%        Indels:          43
DB:                     1             Gaps:            6

US-09-980-881A-2 (1-360) x US-08-696-139-1 (1-1215)

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              |||:::      :::|||                              :::   :::|||
Db         28 TTCGAAGGGGAGAAGGTGTTCCGTGTCAATGTTGAAGATGAAAATGACATCAGCTTACTC 87

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
                    |||   :::|||   :::|||      |||:::|||   :::          |||
Db         88 CATGAGTTGGCCAGCACCAGGCAGATTGACTTCTGGAAACCAGATTCTGTCACACAAATC 147
```

```
Qy          63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                  |||  |||  |||  |||  |||:::    |||:::
Db         148 AAACCTCACAGTACAGTTGACTTCCGTGTGAAAGCAGAAGATATTTTGGCTGTGGAAGAC 207

Qy          83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
                |||     :::  :::       ||||||:::  ::::::     ::::::
Db         208 TTTCTGGAGCAGAATGAACTACAATATGAGGTACTCATAAACAACCTGAGATCTGTGCTC 267

Qy         103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
                :::  |||     :::           |||  :::       ||||||:::
Db         268 GAGGCTCAGTTTGACAGCAGAGTC---------CGTACAACTGGACACAGTTATGAGAAG 318

Qy         123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
                |||::::::       |||  :::|||  :::  :::|||      :::||||||
Db         319 TACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAAAATCCAGAC 378

Qy         143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
                :::::::::::      ||||||::::::|||          :::|||:::||||||
Db         379 CTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATACCTCCTCAAG 438

Qy         163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
                |||  ||||||           |||  ||||||::::::||||||||  ||||||
Db         439 GTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 495

Qy         183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 195
                |||||||||||||  ||||||||  ||||||:::
Db         496 AGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTTCTCACCTAT 555

Qy         195 ---------------------------------------------------------- 195

Db         556 GGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTCTTGCCTGTG 615

Qy         196 ----------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 205
                         |||:::               ||||||||||||||||   |||
Db         616 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCGC 675

Qy         206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
                |||     |||      |||||||||||||||   |||   ||||||
Db         676 TCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 732

Qy         226 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245
                ||||||      |||||||||:::      |||   ||||||||||||||
Db         733 GGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGCT 792

Qy         246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265
                |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db         793 GCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTCC 852

Qy         266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 285
                ||||||||||||:::::::::||||||||||||||   |||:::::||||||||||||
Db         853 ATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTAT 912

Qy         286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 305
                       :::::::::   ||||||:::   :::|||      ||||||:::
Db         913 GATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAAA 972

Qy         306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325
                   :::   |||      |||:::||||||:::|||   |||:::   |||:::
Db         973 GAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAATC 1029

Qy         326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345
                |||  |||  ||||||  |||||||||  ||||||  ||||||||||||||||||
Db        1030 TATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTTC 1089

Qy         346 Thr 346
                |||
Db        1090 ACC 1092

RESULT 7
US-08-860-882A-56
; Sequence 56, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1263 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-56

Alignment Scores:
Pred. No.:                6.34e-70        Length:          1263
Score:                    639.00          Matches:         135
Percent Similarity:       53.83%          Conservative:    69
Best Local Similarity:    35.62%          Mismatches:      131
Query Match:              33.44%          Indels:          44
DB:                       2               Gaps:            7

US-09-980-881A-2 (1-360) x US-08-860-882A-56 (1-1263)

Qy           6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla---PheGln 24
                 |||  |||||||||  :::  |||          |||          |||:::
Db           4 CTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAGCACTTTGAA 63

Qy          25 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 44
                   :::|||                                :::  :::::::::
Db          64 GGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATCCGCGAG 123

Qy          45 LeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLys 64
                 |||  :::|||  :::|||     |||:::|||  :::      |||
Db         124 TTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATCAAACCT 183

Qy          65 LysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 84
                    |||  |||  |||  |||  |||          |||:::      |||
Db         184 CACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAATGTTCTA 243

Qy          85 AsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 104
                   :::  :::       |||||:::::::::::   ::::::::::::
Db         244 AAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTGGAGGCT 303
```

```
Qy        105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHis 124
              |||      :::                |||||:::          |||||:::|||:::
Db        304 CAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAGTACAAC 354

Qy        125 SerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeu 144
                  |||   :::|||   :::   :::          :::|||   ::::::
Db        355 AAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCCCTCATC 414

Qy        145 ThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
              ::::::      ||||||::::::||||||         :::|||:::||||||||||
Db        415 TCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAGGTT--- 471

Qy        165 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGlu 184
              ||||||               |||   ||||||::::::||||||||   ||||||||||
Db        472 GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCCAGAGAG 531

Qy        185 TrpIleSerProAlaPheCysLeuTrpPheIle-------------------------- 195
              ||||||||||||||||||||   ||||||:::
Db        532 TGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTATGGACGT 591

Qy        195 ------------------------------------------------------------ 195

Db        592 GAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTGCTCAAT 651

Qy        196 ------GlyHis-----------------AsnArgMetTrpArgLysAsnArgSerPhe 207
                    |||:::               :::|||   |||||||||   ||||||
Db        652 ATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGCTCCACC 711

Qy        208 TyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
              :::      :::   ||||||||||||||   |||   ||||||           |||
Db        712 CATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCTGGTTGG 768

Qy        228 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 247
              ||||||   |||||||||   :::   |||   ||||||||||||||           |||
Db        769 TGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCCGCAGAG 828

Qy        248 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
              ||||||   |||   ||||||:::|||   |||:::|||     ::::::   ||||||
Db        829 TCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCCATCAAG 888

Qy        268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyrThrArg 287
              ||||||:::::::::|||||||||||||||   :::::::::||||||||||||
Db        889 GCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATATGCTTAC 948

Qy        288 SerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArgAlaIle 307
                  :::::::::   ||||||:::   :::|||          |||:::   :::
Db        949 AAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAAGAACTT 1008

Qy        308 AspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
                  :::         |||:::|||||||:::|||   |||:::   |||:::|||
Db       1009 ---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATCTATCCT 1065

Qy        328 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346
              |||   ||||||   |||||||||   ||||||   ||||||:::||||||||||||
Db       1066 GCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTCACC 1122

RESULT 8
US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 38:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1263 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

Alignment Scores:
Pred. No.:                6.34e-70        Length:          1263
Score:                    639.00          Matches:         135
Percent Similarity:       53.83%          Conservative:    69
Best Local Similarity:    35.62%          Mismatches:      131
Query Match:              33.44%          Indels:          44
DB:                       4               Gaps:            7

US-09-980-881A-2 (1-360) x US-09-011-769A-38 (1-1263)

Qy          6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla---PheGln 24
              |||   |||||||||   :::   |||          |||             |||:::
Db          4 CTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAGCACTTTGAA 63

Qy         25 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 44
                    :::|||                          :::   :::::::::
Db         64 GGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATCCGCGAG 123

Qy         45 LeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLys 64
              |||   :::|||   :::|||       |||:::|||   :::         |||
Db        124 TTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATCAAACCT 183

Qy         65 LysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAlaHisLeu 84
                    |||   |||   |||   |||   |||           |||:::      |||
Db        184 CACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAATGTTCTA 243

Qy         85 AsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGln 104
                    :::   :::          ||||||:::::::::::   ::::::::::::
Db        244 AAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTGGAGGCT 303

Qy        105 GlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHis 124
              |||      :::                |||||:::          |||||:::|||:::
Db        304 CAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAGTACAAC 354

Qy        125 SerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeu 144
                    |||   :::|||   :::   :::          :::|||   ::::::
Db        355 AAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCCCTCATC 414

Qy        145 ThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLysValSer 164
              ::::::      ||||||::::::||||||         :::|||:::||||||||||
```

```
Db        415 TCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAGGTT--- 471

Qy        165 GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGlu 184
              ||||||              |||   ||||||::::::||||||||   ||||||||||
Db        472 GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCCAGAGAG 531

Qy        185 TrpIleSerProAlaPheCysLeuTrpPheIle-------------------------- 195
              ||||||||||||||||||||   ||||||:::
Db        532 TGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTATGGACGT 591

Qy        195 ------------------------------------------------------------ 195

Db        592 GAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTGCTCAAT 651

Qy        196 ------GlyHis-----------------AsnArgMetTrpArgLysAsnArgSerPhe 207
                     |||:::                ::::|||   |||||||||   ||||||
Db        652 ATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGCTCCACC 711

Qy        208 TyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrp 227
              :::      :::  ||||||||||||||   |||   ||||||            |||
Db        712 CATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCTGGTTGG 768

Qy        228 CysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGlu 247
              ||||||   |||||||||   :::   |||   ||||||||||||||          |||
Db        769 TGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCCGCAGAG 828

Qy        248 SerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLys 267
              ||||||   |||   ||||||:::|||   |||:::|||      ::::::   ||||||
Db        829 TCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCCATCAAG 888

Qy        268 AlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyrThrArg 287
              ||||||:::::::::|||||||||||||||   :::::::::||||||||||||
Db        889 GCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATATGCTTAC 948

Qy        288 SerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArgAlaIle 307
                  :::::::::   ||||||:::   :::|||         |||:::   :::
Db        949 AAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAAGAACTT 1008

Qy        308 AspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeu 327
                  :::         |||:::||||||:::|||   |||:::   |||:::|||
Db       1009 ---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATCTATCCT 1065

Qy        328 AlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPheThr 346
              |||   ||||||   ||||||||   ||||||   ||||||:::||||||||||||
Db       1066 GCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTCACC 1122

RESULT 9
US-08-860-882A-71
; Sequence 71, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1284 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-71

Alignment Scores:
Pred. No.:              5.67e-69        Length:         1284
Score:                  631.50          Matches:        129
Percent Similarity:     54.57%          Conservative:   68
Best Local Similarity:  35.73%          Mismatches:     121
Query Match:            33.05%          Indels:         43
DB:                     2               Gaps:           6

US-09-980-881A-2 (1-360) x US-08-860-882A-71 (1-1284)

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              |||:::      :::|||                                 :::   ::::::
Db         85 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 144

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              :::   |||   :::|||   :::|||      |||:::|||  :::        |||
Db        145 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 204

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                        |||   |||   |||   |||   |||         |||:::
Db        205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 264

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
                 |||       :::   :::         |||||:::::::::::::  :::::::::
Db        265 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
              :::   |||      :::          ||||||:::          ||||||:::
Db        325 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 375

Qy        123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
              |||:::            |||   :::|||  :::   :::         :::|||
Db        376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 435

Qy        143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
              :::::::::::       ||||||::::::||||||        :::|||:::||||||
Db        436 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 495

Qy        163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
              |||   ||||||            |||   ||||||:::::||||||||   ||||||
Db        496 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552

Qy        183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle-------------------- 195
              ||||||||||||||||||||||||||   ||||||:::
Db        553 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 612

Qy        195 ------------------------------------------------------------ 195

Db        613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 672

Qy        196 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 205
```

```
Db        673 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 732

Qy        206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225

Db        733 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy        226 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245

Db        790 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849

Qy        246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265

Db        850 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 909

Qy        266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 285

Db        910 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 969

Qy        286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 305

Db        970 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy        306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325

Db       1030 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1086

Qy        326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345

Db       1087 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146

Qy        346 Thr 346

Db       1147 ACC 1149

RESULT 10
US-09-011-769A-55
; Sequence 55, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 55:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1284 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION: 1..1272
;        FEATURE:
;             NAME/KEY:  mat_peptide
;             LOCATION: 352..1272
;        SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-011-769A-55

Alignment Scores:
Pred. No.:              5.67e-69      Length:          1284
Score:                  631.50        Matches:         129
Percent Similarity:     54.57%        Conservative:    68
Best Local Similarity:  35.73%        Mismatches:      121
Query Match:            33.05%        Indels:          43
DB:                     4             Gaps:            6

US-09-980-881A-2 (1-360) x US-09-011-769A-55 (1-1284)

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42

Db         85 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 144

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62

Db        145 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 204

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82

Db        205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 264

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102

Db        265 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122

Db        325 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 375

Qy        123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142

Db        376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 435

Qy        143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162

Db        436 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 495

Qy        163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182

Db        496 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552

Qy        183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 195

Db        553 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 612

Qy        195 ---------- 195

Db        613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 672

Qy        196 ----------GlyHis------------------AsnArgMetTrpArgLysAsnArg 205

Db        673 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 732

Qy        206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
```

```
Db        733 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy        226 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 245

Db        790 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849

Qy        246 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 265

Db        850 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 909

Qy        266 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 285

Db        910 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 969

Qy        286 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 305

Db        970 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy        306 AlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 325

Db       1030 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1086

Qy        326 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 345

Db       1087 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146

Qy        346 Thr 346

Db       1147 ACC 1149
```

RESULT 11
US-09-171-945-124
; Sequence 124, Application US/09171945
; Patent No. 6277599
; GENERAL INFORMATION:
; APPLICANT: Emery, Stephen
; APPLICANT: Copley, Clive Graham
; APPLICANT: Edge, Michael Derek
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
; TITLE OF INVENTION: Antibody, and Their Therapeutic Use in an Adept System
; FILE REFERENCE: Monoclonal Antibody to CEA
; CURRENT APPLICATION NUMBER: US/09/171,945
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: GB9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; NUMBER OF SEQ ID NOS: 131
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 124
; LENGTH: 2154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: humanized .
US-09-171-945-124

Alignment Scores:

Pred. No.:	1.59e-67	Length:	2154
Score:	623.00	Matches:	132
Percent Similarity:	53.40%	Conservative:	72
Best Local Similarity:	34.55%	Mismatches:	134
Query Match:	32.60%	Indels:	44
DB:	3	Gaps:	7

US-09-980-881A-2 (1-360) x US-09-171-945-124 (1-2154)

```
Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22

Db          1 ATGTTGGCACTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAG 60

Qy         23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41

Db         61 CACTTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATA 120

Qy         42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61

Db        121 ATCCGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAA 180

Qy         62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81

Db        181 ATCAAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAG 240

Qy         82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101

Db        241 AATGTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTG 300

Qy        102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121

Db        301 GTGGAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAG 351

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141

Db        352 AAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCA 411

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161

Db        412 GCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTG 471

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181

Db        472 AAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCAT 528

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 195

Db        529 GCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACC 588

Qy        195 ---------- 195

Db        589 TATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCT 648

Qy        196 ----------GlyHis----------AsnArgMetTrpArgLysAsn 204

Db        649 GTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACT 708

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224

Db        709 CGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GAT 765

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeu 244

Db        766 GCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCT 825

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264

Db        826 GCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCT 885

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284

Db        886 TCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCA 945

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304

Db        946 TATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTG 1005

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324

Db       1006 AAAGAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACA 1062

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344

Db       1063 ATCTATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTATGACCAAGGAATCAGATATTCC 1122

Qy        345 PheThr 346
```

```
                ||||||
Db       1123 TTCACC 1128

RESULT 12
US-09-675-305-9
; Sequence 9, Application US/09675305
; Patent No. 6441153
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. 6441153el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/09/675,305
;  CURRENT FILING DATE:  2000-09-29
;  PRIOR APPLICATION NUMBER: US 60/156,685
;  PRIOR FILING DATE: 1999-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-09-675-305-9

Alignment Scores:
Pred. No.:                1.89e-66      Length:          1311
Score:                    611.50        Matches:         141
Percent Similarity:       50.36%        Conservative:    69
Best Local Similarity:    33.81%        Mismatches:      135
Query Match:              32.00%        Indels:          72
DB:                       4             Gaps:            11

US-09-980-881A-2 (1-360) x US-09-675-305-9 (1-1311)

Qy          1 MetLysLeuCys------SerLeuAlaValLeuValProIleValLeuPheCysGluGln 18
              :::   ||||||         |||   :::|||   |||
Db         43 CTGCCTCTTTGCTGGCTCTTTTTGAAGATTCTGCAACCG-----------GGGCACAGC 90

Qy         19 HisValPhe------AlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSer 36
              |||::::::        :::         :::|||:::      :::|||:::|||
Db         91 CACCTTTATAACAACCGCTATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAA 150

Qy         37 ArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnPro 56
                 :::        |||:::   ::::::        ::::::   ||||||||||||
Db        151 GAGGAAGCATATGCACTGAAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCC 210

Qy         57 ValThrAlaAspLeuIleValLys------LysLysGlnValHisPheValAsnAla 74
                 :::        :::   :::              ||||||         |||
Db        211 AGCAGTATCTCCTATGTATCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGT 270

Qy         75 SerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeu 94
              |||           :::   |||   |||     :::   |||         ||||||
Db        271 TCCCGAGCC------CTGTTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTC 324

Qy         95 LeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrValSerProArg 114
              :::   |||::::::      :::::::::      |||       |||         |||
Db        325 ATAGAAGATCTTCAGAAAACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGA 384

Qy        115 AlaSerAlaSer------TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrp 132
              |||   |||         ||||||   ||||||||||||   ||||||   :::|||
Db        385 AGATCCCTCTCTGGATATAATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGG 444

Qy        133 IleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSer 152
              :::     :::   :::   |||     ::::::              ||||||   |||
Db        445 ATGCATCATCTGAATAAAACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCA 504

Qy        153 PheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 172
              :::|||         |||::::::||||||:::   |||:::            |||

Db        505 TATGAGGGAAGATCTCTTTTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGA 561

Qy        173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeu 192
              |||:::||||||||||||||||||||||||||||||||||   ||||||||||||
Db        562 GCTGTTTGGATAGACTGTGGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAG 621

Qy        193 TrpPheIle---------------------------------------------------- 195
              ||||||:::
Db        622 TGGTTTGTAAAAGAAGCTCTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTG 681

Qy        195 ---------------------------------------------------------- 195

Db        682 AATCATCTATATTTCTATATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGG 741

Qy        196 GlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGly 215
                 ::::::|||   |||||||||   ||||||      :::         |||   |||
Db        742 ACCAATGATCGATTTTGGAGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGA 801

Qy        216 ThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSer 235
                 |||   |||   |||:::      |||   |||||||:::||||||||||||
Db        802 GTGGATGCCAATAGAAACTGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCAC 858

Qy        236 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaVal 255
                 |||   :::|||||||||||   :::||||||||||||||||||||||||||||||
Db        859 CCTTGTGATGACACATACTGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTA 918

Qy        256 AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyr 275
                 |||:::|||||||||::::::         |||:::||||||:::|||   |||:::|||
Db        919 GCTAACTTCCTTCGAAAACACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATAT 978

Qy        276 SerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGlu 295
                 :::|||   :::::::::||||||||||||      :::         :::
Db        979 GCTCAGATGTTACTGTATCCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGT 1038

Qy        296 LeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArg 315
                 :::         |||   :::||||||   |||:::   ::::::              |||
Db       1039 GTGGAATCTGCAGCTTATAAAGCTGTGAATGCACTT---CAGTCAGTATACGGGGTACGA 1095

Qy        316 TyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrp 335
                 |||   :::|||       |||   |||||||||::::::   |||         ||||||
Db       1096 TACAGATATGGACCAGCCTCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGG 1155

Qy        336 IleTyrAspLeuGlyIleLysTyrSerPhe---------------------------- 345
                 |||       ||||||   |||:::|||
Db       1156 GCCTACAAAAATGGAATACCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTT 1215

Qy        346 ----------------------ThrSerAsnProProValGluLysLeu 354
                                     :::||||||||||||||:::||||||
Db       1216 GGATTTTTACTCCCAGAGATGCTCATCAAACCCACCTGTACAGAAACTA 1264

RESULT 13
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
;  GENERAL INFORMATION:
;    APPLICANT:  Hartman, Jacob
;    APPLICANT:  Fulga, Netta
;    APPLICANT:  Mendelovitch, Simona
;    APPLICANT:  Gorecki, Marian
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,760
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/378,233
;      FILING DATE:  25-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
US-08-782-760-5

Alignment Scores:
Pred. No.:              9.11e-66      Length:        927
Score:                  604.00        Matches:       125
Percent Similarity:     60.14%        Conservative:  44
Best Local Similarity:  44.48%        Mismatches:    72
Query Match:            31.61%        Indels:        41
DB:                     2             Gaps:          5

US-09-980-881A-2 (1-360) x US-08-782-760-5 (1-927)

Qy        115 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 134
              ||||||          |||   :::|||::::::          |||   :::|||||::: 
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy        135 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
                :::         :::||||||::::::|||:::       ||||||::::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy        155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 174
                  :::||||||||||:::   ||||||   :::       |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy        175 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 194
              :::||||||||||||   ||||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy        195 Ile---------------------------------------------------------- 195
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297

Qy        196 ----------------------------------GlyHis------------------ 197
                                                 |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy        198 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 217
              :::||||||||||||||   ||||||      |||   :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 237
                  |||   ||||||   :::      |||||||||   |||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        238 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 257
              ||||||||||||||||||      ||||||||||||   |||   ||||||:::|||

Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        258 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
              |||:::|||   |||::::::   ||||||||||||:::::::::||||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 297
                  :::::::::|||||||||||                :::::::::|||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        298 LeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThr 317
                  :::         |||   :::   :::       |||          |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        318 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 337
                  :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        338 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 357
              |||   ||||||||||||||||||||          |||         :::||||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        358 Ser 358
              |||
Db        831 TCT 833

RESULT 14
PCT-US96-00995-5
; Sequence 5, Application PC/TUS9600995
;  GENERAL INFORMATION:
;    APPLICANT:  Bio-Technology General Corp.
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00995
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847-A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
PCT-US96-00995-5

Alignment Scores:
Pred. No.:              9.11e-66      Length:        927
```

```
Score:                 604.00        Matches:       125
Percent Similarity:    60.14%        Conservative:   44
Best Local Similarity: 44.48%        Mismatches:     72
Query Match:           31.61%        Indels:         41
DB:                    5             Gaps:           5

US-09-980-881A-2 (1-360) x PCT-US96-00995-5 (1-927)

Qy        115 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 134
              ||||||         |||   :::|||::::::         |||   :::||||||:::
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy        135 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 154
                 :::         :::||||||::::::|||:::      ||||||::::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy        155 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 174
                    :::||||||||||||:::   ||||||   :::      |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy        175 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 194
              :::||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy        195 Ile---------------------------------------------------------- 195
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297

Qy        196 ----------------------------------GlyHis------------------ 197
                                                 |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy        198 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 217
              :::||||||||||||||   ||||||      |||  :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        218 LeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 237
                 |||   ||||||   :::      |||||||||   |||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        238 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 257
              ||||||||||||||||      ||||||||||||   |||   ||||||:::|||
Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        258 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 277
              |||:::|||   |||::::::   |||||||||||:::::::::||||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        278 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 297
                 :::::::::|||||||||||||               :::::::::||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        298 LeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThr 317
                 :::          |||   :::   :::      |||          |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        318 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 337
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        338 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 357
              |||   ||||||||||||||||||||         |||      :::||||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        358 Ser 358
              |||
Db        831 TCT 833

RESULT 15
US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..921
US-08-696-139-3

Alignment Scores:
Pred. No.:             2.9e-62       Length:        921
Score:                 576.00        Matches:       116
Percent Similarity:    59.09%        Conservative:   40
Best Local Similarity: 43.94%        Mismatches:     68
Query Match:           30.14%        Indels:         40
DB:                    1             Gaps:           5

US-09-980-881A-2 (1-360) x US-08-696-139-3 (1-921)

Qy        120 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 139
              ||||||:::|||::::::         |||   :::|||   :::   :::|||
Db         16 TATGAGAAGTACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAA 75

Qy        140 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 159
                 :::||||||:::::::::::      ||||||::::::|||            :::|||
Db         76 AATCCAGACCTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATAC 135

Qy        160 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 179
                 :::||||||||||   ||||||           |||   ||||||::::::||||||||
Db        136 CTCCTCAAGGTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGT 192

Qy        180 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 195
                   ||||||||||||||||||||   |||||||||   |||||||:::
Db        193 TTCCATGCCAGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTT 252
```

```
Qy      195 ------------------------------------------------------------ 195

Db      253 CTCACCTATGGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTC 312

Qy      196 --------------------GlyHis-----------------AsnArgMetTrpArg 202
                               |||:::                 ||||||||||||||
Db      313 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 372

Qy      203 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPhe 222
            |||   ||||||      |||          ||||||||||||||   |||   ||||||
Db      373 AAGACCCGCTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT 432

Qy      223 ValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 242
                  ||||||      |||||||||:::       |||   ||||||||||||
Db      433 ---GATGCTGGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGT 489

Qy      243 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 262
            |||         |||||||||   |||   ||||||:::|||   |||:::|||   |||
Db      490 GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAAC 549

Qy      263 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 282
            ::::::   ||||||||||||:::::::::|||||||||||||||   |||::::::|||
Db      550 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCT 609

Qy      283 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 302
            |||||||||                 :::::::::   ||||||:::   :::|||
Db      610 TATTCCTATGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCT 669

Qy      303 AlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 322
            ||||||:::   :::      |||          |||:::||||||:::|||   |||:::
Db      670 GCCGTGAAAGAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCT 726

Qy      323 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 342
               |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db      727 ACAACAATCTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 786

Qy      343 TyrSerPheThr 346
            ||||||||||||
Db      787 TATTCCTTCACC 798

Search completed: January  2, 2004, 09:25:32
Job time : 193.344 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 20:10:07 ; Search time 875.574 Seconds
(without alignments)
1423.134 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135919_26949/app_query.fasta_1.1628
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980881_@CGN_1_1_704_@runat_31122003_135919_26949
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1832	95.9	1728	10	US-09-880-107-2396	Sequence 2396, Ap
2	1729	90.5	1272	13	US-10-379-836-1	Sequence 1, Appli
3	889	46.5	1400	9	US-09-925-302-24	Sequence 24, Appl
4	623	32.6	2154	10	US-09-910-059-124	Sequence 124, App
5	621.5	32.5	1332	10	US-09-954-456-1141	Sequence 1141, Ap
6	620.5	32.5	1633	13	US-10-341-434-187	Sequence 187, App
7	619.5	32.4	1254	15	US-10-229-546-3	Sequence 3, Appli
8	619.5	32.4	1622	12	US-10-429-802-21	Sequence 21, Appl
9	619.5	32.4	1622	15	US-10-229-546-1	Sequence 1, Appli
10	611.5	32.0	1311	14	US-10-200-344-9	Sequence 9, Appli
11	611.5	32.0	1993	12	US-10-274-639-33	Sequence 33, Appl
12	591	30.9	416	10	US-09-960-352-14595	Sequence 14595, A
13	564	29.5	2128	14	US-10-200-344-13	Sequence 13, Appl
14	563	29.5	1125	9	US-09-888-615-2	Sequence 2, Appli
15	563	29.5	1332	16	US-10-176-306-75	Sequence 75, Appl
16	563	29.5	1603	16	US-10-176-306-73	Sequence 73, Appl
17	556	29.1	1826	13	US-10-252-157-453	Sequence 453, App
18	546	28.6	1200	15	US-10-200-910-7	Sequence 7, Appli
19	543	28.4	1870	10	US-09-910-059-112	Sequence 112, App
20	528.5	27.7	1050	14	US-10-200-344-11	Sequence 11, Appl
21	519.5	27.2	1311	15	US-10-200-910-5	Sequence 5, Appli
22	511.5	26.8	1311	14	US-10-076-535-3	Sequence 3, Appli
23	511.5	26.8	1687	14	US-10-076-535-1	Sequence 1, Appli
24	508	26.6	1306	9	US-09-923-779-144	Sequence 144, App
25	508	26.6	1306	9	US-09-962-436-557	Sequence 557, App
26	504	26.4	1341	9	US-09-925-297-69	Sequence 69, Appl
27	489.5	25.6	1254	9	US-09-925-297-58	Sequence 58, Appl
28	489.5	25.6	1260	13	US-10-345-680-66	Sequence 66, Appl
29	489.5	25.6	1380	10	US-09-969-347-209	Sequence 209, App
30	489.5	25.6	1380	13	US-10-345-680-64	Sequence 64, Appl
31	487	25.5	397	10	US-09-960-352-8425	Sequence 8425, Ap
32	483	25.3	2786	11	US-09-946-374-233	Sequence 233, App
33	483	25.3	2786	13	US-10-015-387A-233	Sequence 233, App
34	483	25.3	2786	13	US-10-006-130A-233	Sequence 233, App
35	483	25.3	2786	13	US-10-199-672-307	Sequence 307, App
36	483	25.3	2786	13	US-10-006-172A-233	Sequence 233, App
37	483	25.3	2786	13	US-10-187-749-307	Sequence 307, App
38	483	25.3	2786	13	US-10-194-457-307	Sequence 307, App
39	483	25.3	2786	13	US-10-184-642-307	Sequence 307, App
40	483	25.3	2786	13	US-10-196-747-307	Sequence 307, App
41	483	25.3	2786	13	US-10-015-392A-233	Sequence 233, App
42	483	25.3	2786	13	US-10-017-253A-233	Sequence 233, App
43	483	25.3	2786	13	US-10-173-689-307	Sequence 307, App
44	483	25.3	2786	13	US-10-173-690-307	Sequence 307, App
45	483	25.3	2786	13	US-10-173-691-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Alignment Scores:
Pred. No.:              3.67e-222       Length:         1728
Score:                  1832.00         Matches:        357
Percent Similarity:     86.27%          Conservative:   1
Best Local Similarity:  86.02%          Mismatches:     2
Query Match:            95.87%          Indels:         55
DB:                     10              Gaps:           2

US-09-980-881A-2 (1-360) x US-09-880-107-2396 (1-1728)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         20 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 79

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 139

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        500 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATC 559

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        560 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 619

Qy        197 ------------------------------------------------------------ 197

Db        620 TTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATG 679

Qy        198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db        680 CCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 739

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        740 AACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCT 799

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        800 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGA 859

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        860 CTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATC 919

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        920 AACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATAT 979

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        980 TCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCA 1039

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db       1040 GTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAA 1099

Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1100 ACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATAT 1159

Qy        344 SerPhe----------------------------------------------------Th 346
              ||||||                                                    ||
Db       1160 TCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTAC 1219

Qy        346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
              |||||||||||||||||||||||||||||||||||||||||||
Db       1220 ATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAA 1262

RESULT 2
US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1272
;   TYPE: DNA
;   ORGANISM: Papio hamadryas
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1269)
US-10-379-836-1

Alignment Scores:
Pred. No.:              2.8e-209        Length:         1272
Score:                  1729.00         Matches:        338
Percent Similarity:     83.37%          Conservative:   8
Best Local Similarity:  81.45%          Mismatches:     14
Query Match:            90.48%          Indels:         55
DB:                     13              Gaps:           2

US-09-980-881A-2 (1-360) x US-10-379-836-1 (1-1272)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAGCTTTGCAGTCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 60

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTCGCGTTTCAGAGTGGCCAGGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 120

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GTGCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCGGAC 180

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||  ||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        181 CTTATTGAGAAGAAAAAACAAGTCCATTTTTTTGTAAATTCATCTGATGTCGACAATGTG 240
```

```
Qy        81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCCTGCTGGCAGATGTGGAAGAT 300

Qy       101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 360

Qy       121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
             ||||||||||||||||||||||||||||||||||||||    |||||||||::::::
Db       361 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAACTTATAACTGAGAAGTAT 420

Qy       141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
             ||||||||||||||||||||||||||||||||:::||||||:::||||||||||||||
Db       421 CCTGATATGCTTACAAAAATCCACATTGGATCCTCCTATGAGAAGCACCCACTTTATGTT 480

Qy       161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
             ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db       481 TTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATGTGGATTGACTGTGGAATC 540

Qy       181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CATGCCAGAGAATGGATCTCCCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTGAA 600

Qy       197 ------------------------------------------------------------ 197

Db       601 TACTACGGGATAATAGGGGAATATACCAATCTTCTGAGGCATGTGGATTTCTATGTTATG 660

Qy       198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                     ||||||||||||||||
Db       661 CCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAG 720

Qy       204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
             ||||||||||||||||||||||   ||||||||||||||||||||   ||||||
Db       721 AACCGTTCTTTCTATGCGAACAATCGTTGCATCGGAACAGACCTGAACAGGAACTTTGCG 780

Qy       224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
             ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db       781 TCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTTCTCATGCTCGGAAACCTACTGTGGA 840

Qy       244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
             ||||||||||||||||||||||   ||||||||||:::||||||||||||||||||||||
Db       841 CTTTATCCTGAGTCAGAACCAGAAGCGAAGGCGGTGGCTAATTTCTTGAGAAGAAATATC 900

Qy       264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
             |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 AACCACATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATCGTGTTTCCATAT 960

Qy       284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TCCTATACTCGAAGCAAAAGCAAAGACCACGAGGAATTGTCTCTAGTAGCCAGTGAAGCA 1020

Qy       304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
             ||||||||||||   ||||||||||||||   ||||||||||||||   |||||||||
Db      1021 GTTCGTGCTATTCAGAAAACCAGTAAAAATATCAGGTATACACATGGCCGTGGCTCAGAA 1080

Qy       324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
             ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db      1081 ACCTTATACCTAGCTCCTGGAGGTGCGGACGATTGGATCTATGATTTGGGCATCAAATAT 1140

Qy       344 SerPhe-------------------------------------------------------Th 346
             ||||||                                                       ||
Db      1141 TCGTTTACAATTGAACTTCGAGATACGGGCAAATACGGATTCTTGCTGCCTGAGCGTTAC 1200

Qy       346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
             ||||||||||   |||:::   ||||||||||||||||||||
Db      1201 ATCAAACCCACTTGTAAAGACGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 3
US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1

; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (25)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Alignment Scores:
Pred. No.:              2.58e-102     Length:          1400
Score:                  889.00        Matches:         178
Percent Similarity:     75.21%        Conservative:    1
Best Local Similarity:  74.79%        Mismatches:      3
Query Match:            46.52%        Indels:          56
DB:                     9             Gaps:            2

US-09-980-881A-2 (1-360) x US-09-925-302-24 (1-1400)

Qy       179 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 197
             ||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        10 GGAATCCATGCCAGANAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATA 69

Qy       197 ------------------------------------------------------------ 197

Db        70 ACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTAT 129

Qy       198 ----------------------------------------------AsnArgMetTr 201
                                                           ||||||||||
Db       130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACTACTCATGGAAAAAGAATCGAATGTG 189

Qy       201 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAs 221
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||   ||
Db       190 GAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAA 249

Qy       221 nPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 241
             ||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       250 CTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309

Qy       241 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       310 CTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAG 369

Qy       261 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 281
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       370 AAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTT 429

Qy       281 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 301
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       430 TCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAG 489

Qy       301 rGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 321
             ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||
Db       490 TGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGG 549

Qy       321 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIl 341
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609

Qy       341 eLysTyrSerPhe------------------------------------------- 345
```

```
              ||||||||||||
Db        610 CAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGA 669

Qy        346 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
                   |||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 719

RESULT 4
US-09-910-059-124
; Sequence 124, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 124
;   LENGTH: 2154
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-124

Alignment Scores:
Pred. No.:                3.77e-68        Length:          2154
Score:                    623.00          Matches:         132
Percent Similarity:       53.40%          Conservative:    72
Best Local Similarity:    34.55%          Mismatches:      134
Query Match:              32.60%          Indels:          44
DB:                       10              Gaps:            7

US-09-980-881A-2 (1-360) x US-09-910-059-124 (1-2154)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              :::  :::|||  |||||||||  :::  |||           |||
Db          1 ATGTTGGCACTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAG 60

Qy         23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
                |||:::     :::|||                :::   :::
Db         61 CACTTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATA 120

Qy         42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61
              ::::::  |||  :::|||  :::|||      |||:::|||  :::
Db        121 ATCCGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAA 180

Qy         62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
              |||                |||  |||  |||  |||  |||         |||:::
Db        181 ATCAAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAG 240

Qy         82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
                |||     :::  :::      ||||||::::::::::::   ::::::
Db        241 AATGTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTG 300

Qy        102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ::::::  |||      ::               ||||||:::         ||||||
Db        301 GTGGAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAG 351

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              :::|||:::           |||  :::|||  :::  :::        :::|||
Db        352 AAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCA 411

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
              ::::::::::::      ||||||::::::||||||        :::|||:::|||
Db        412 GCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTG 471

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              ||||||  ||||||         |||  ||||||::::::|||||||   |||
Db        472 AAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCAT 528

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 195
              ||||||||||||||||||||||||||||  ||||||::
Db        529 GCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACC 588

Qy        195 ---------------------------------------------------- 195

Db        589 TATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCT 648

Qy        196 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204
                           |||:::                 :::|||  |||||||||
Db        649 GTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACT 708

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||   :::      :::  ||||||||||||||   |||  ||||||
Db        709 CGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GAT 765

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
                  |||||||||  |||||||||  :::  |||  ||||||||||||||||
Db        766 GCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCT 825

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
                  |||||||||  |||  ||||||:::|||  |||:::|||      ::::::
Db        826 GCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCT 885

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284
              ||||||||||||:::::::::||||||||||||||||   :::::::::||||||||
Db        886 TCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCA 945

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              |||                :::::::::  ||||||:::  :::|||         |||
Db        946 TATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTG 1005

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
              :::  :::      :::      |||:::||||||:::|||  |||:::  |||
Db       1006 AAAGAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACA 1062

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              :::|||  :::  |||        ||||||  ||||||  ||||||:::|||||||
Db       1063 ATCTATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTATGACCAAGGAATCAGATATTCC 1122

Qy        345 PheThr 346
              ||||||
Db       1123 TTCACC 1128

RESULT 5
US-09-954-456-1141
; Sequence 1141, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
```

```
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1141
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-1141

Alignment Scores:
Pred. No.:                2.6e-68        Length:          1332
Score:                    621.50         Matches:         134
Percent Similarity:       53.66%         Conservative:    71
Best Local Similarity:    35.08%         Mismatches:      132
Query Match:              32.52%         Indels:          45
DB:                       10             Gaps:            8

US-09-980-881A-2 (1-360) x US-09-954-456-1141 (1-1332)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              :::   :::|||   |||||||||   :::   |||             |||
Db         24 ATGTTGGCACTCTTGGTTCTGGTGACTGTGGCCCTGGCATCTGCTCATCATGGTGGTGAG 83

Qy         23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
                 |||:::   :::|||                                 :::  :::
Db         84 CACTTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATA 143

Qy         42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61
              ::::::   |||   :::|||   :::|||     |||:::|||   :::
Db        144 ATCCGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAA 203

Qy         62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
              |||                 |||   |||   |||   |||   |||          |||:::
Db        204 ATCAAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAG 263

Qy         82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
                 |||      :::  :::         ||||||:::::::::::   ::::::
Db        264 AATGTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTG 323

Qy        102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ::::::   |||     :::             ||||||:::          ||||||
Db        324 GTGGAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAG 374

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              :::|||:::             |||   :::|||   :::   :::         :::|||
Db        375 AAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCA 434

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
                 ::::::::::::       ||||||::::::||||||        :::|||:::|||
Db        435 GCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTG 494

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              ||||||   ||||||              |||   ||||||::::::|||||||||   |||
Db        495 AAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCAT 551

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 195
              ||||||||||||||||||||||||||||||  ||||||::
Db        552 GCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACC 611

Qy        195 ------------------------------------------------------------ 195
```

```
Db        612 TATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCT 671

Qy        196 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204
                           |||:::                 :::|||   |||||||||
Db        672 GTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACT 731

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||   :::       :::      ||||||||||||   |||   ||||||
Db        732 CGCTCCACCCATACTGGATCT---AGCATTGGCACAGACCCCAACAGAAATTTT---GAT 785

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
                  |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db        786 GCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCT 845

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
                  |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db        846 GCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCT 905

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284
              ||||||||||||:::::::::|||||||||||||||   :::::::::|||||||||
Db        906 TCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCA 965

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              |||                 :::::::::   ||||||:::   :::|||         |||
Db        966 TATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTG 1025

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
              :::   :::      :::          |||:::||||||:::|||   |||:::   |||
Db       1026 AAAGAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACA 1082

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              :::|||   |||   ||||||   |||||||||   ||||||   ||||||:::||||||
Db       1083 ATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCC 1142

Qy        345 PheThr 346
              ||||||
Db       1143 TTCACC 1148

RESULT 6
US-10-341-434-187
; Sequence 187, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
;   LENGTH: 1633
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12)..(1262)
;   OTHER INFORMATION:
US-10-341-434-187

Alignment Scores:
Pred. No.:                4.92e-68       Length:          1633
Score:                    620.50         Matches:         130
Percent Similarity:       54.19%         Conservative:    77
Best Local Similarity:    34.03%         Mismatches:      128
Query Match:              32.47%         Indels:          47
DB:                       13             Gaps:            8

US-09-980-881A-2 (1-360) x US-10-341-434-187 (1-1633)
```

```
Qy          5 SerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla------ 22
              ::::::   :::::::::|||:::   |||                        |||
Db          9 ACCATGAGGCTCATCCTGCCTGTGGGTTTGATTGCTACCACTCTTGCAATTGCTCCTGTC 68

Qy         23 ---PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
                 |||         :::|||         |||:::      :::|||      :::
Db         69 CGCTTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATC 128

Qy         42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61
              ::::::::::|||      |||   |||:::      |||   |||
Db        129 ATAAAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCAC 188

Qy         62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
              :::                |||   |||   |||:::      :::         ::::::
Db        189 GTAGCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAG 248

Qy         82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
              :::   |||:::   :::   :::         :::|||:::   |||:::::::::
Db        249 TCTGCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAG 308

Qy        102 IleGlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              |||::::::|||           :::   :::     |||   |||         |||
Db        309 ATTGAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TAC 359

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
                 :::|||::::::      :::|||   :::|||   |||   :::   :::::::::
Db        360 GCAAAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTAT 419

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              |||:::|||::::::::::|||   |||||||||:::   |||       ||||||||||||
Db        420 CCTGAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTT 479

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||:::   |||::::::         :::   ||||||::::::|||||||||||||
Db        480 CTGAAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATT 536

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle--------------- 195
              |||||||||||||||:::|||||||||||||||   ||||||:::
Db        537 CACGCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAA 596

Qy        195 ------------------------------------------------------------ 195

Db        597 ACTTATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTT 656

Qy        196 ------------------GlyHis------------------AsnArgMetTrpArgLys 203
                               |||:::                  ||||||||||||||||||
Db        657 CCTGTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAA 716

Qy        204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
              |||||||||         |||:::   ||||||||||||||||||||   ||||||
Db        717 AATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT--- 773

Qy        224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
              :::      |||                 ::::::      |||::::::   |||   |||
Db        774 AATGCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGC 833

Qy        244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
                       ||||||||||||   |||   |||||||||   :::|||:::|||   ::::::
Db        834 TCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTG 893

Qy        264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyr 283
              |||:::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        894 AATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTAT 953

Qy        284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
                 ||||||                 :::||||||:::|||:::   ||||||
Db        954 GGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGC 1013

Qy        304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
                    :::      |||         |||||||||   :::|||
Db       1014 ACTGATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCA 1070
```

```
Qy        324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
              |||:::|||          |||          ||||||   ||||||||||||||||:::
Db       1071 ACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACAC 1130

Qy        344 SerPhe 345
              :::|||
Db       1131 ACATTT 1136
```

RESULT 7
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MPI01-156P1RNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-3

Alignment Scores:

Pred. No.:	4.22e-68	Length:	1254
Score:	619.50	Matches:	130
Percent Similarity:	54.07%	Conservative:	76
Best Local Similarity:	34.12%	Mismatches:	128
Query Match:	32.42%	Indels:	47
DB:	15	Gaps:	8

US-09-980-881A-2 (1-360) x US-10-229-546-3 (1-1254)

```
Qy          6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla--------- 22
              :::   :::::::::|||:::   |||                        |||
Db          1 ATGAGGCTCATCCTGCCTGTGGGTTTGATTGCTACCACTCTTGCAATTGCTCCTGTCCGC 60

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              |||         :::|||         |||:::      :::|||      ::::::
Db         61 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 120

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              ::::::|||      |||   |||:::      |||   |||            :::
Db        121 AAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTA 180

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                        |||   |||   |||:::      :::         :::::::::
Db        181 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 240

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
                 |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db        241 GCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 300

Qy        103 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ::::::|||           :::   :::     |||   |||         |||
Db        301 GAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCA 351

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
                 :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||
Db        352 AAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCT 411

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
```

```
Db        412 GAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTG 471

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181

Db        472 AAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCAC 528

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----------------- 195

Db        529 GCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACT 588

Qy        195 ------------------------------------------------------------ 195

Db        589 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 648

Qy        196 ---------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204

Db        649 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 708

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224

Db        709 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 765

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244

Db        766 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264

Db        826 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 885

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 284

Db        886 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 945

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304

Db        946 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1005

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324

Db       1006 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1062

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344

Db       1063 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1122

Qy        345 Phe 345

Db       1123 TTT 1125
```

RESULT 8
US-10-429-802-21
; Sequence 21, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-429-802-21

Alignment Scores:

Pred. No.:	6.51e-68	Length:	1622
Score:	619.50	Matches:	130
Percent Similarity:	54.07%	Conservative:	76
Best Local Similarity:	34.12%	Mismatches:	128
Query Match:	32.42%	Indels:	47
DB:	12	Gaps:	8

US-09-980-881A-2 (1-360) x US-10-429-802-21 (1-1622)

```
Qy          6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla--------- 22

Db          1 ATGAGGCTCATCCTGCCTGTGGGTTTGATTGCTACCACTCTTGCAATTGCTCCTGTCCGC 60

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42

Db         61 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 120

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62

Db        121 AAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTA 180

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82

Db        181 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 240

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102

Db        241 GCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 300

Qy        103 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121

Db        301 GAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCA 351

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141

Db        352 AAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCT 411

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161

Db        412 GAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTG 471

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181

Db        472 AAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCAC 528

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----------------- 195

Db        529 GCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACT 588

Qy        195 ------------------------------------------------------------ 195

Db        589 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 648

Qy        196 ---------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204

Db        649 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 708

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224

Db        709 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 765

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244

Db        766 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264

Db        826 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 885

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 284

Db        886 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 945
```

```
Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              ||||||                 :::|||||||:::|||:::   ||||||
Db        946 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1005

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
                  :::     |||          |||||||||   :::|||            |||
Db       1006 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1062

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              :::|||          |||          ||||||   ||||||||||||||||||:::::
Db       1063 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1122

Qy        345 Phe 345
              |||
Db       1123 TTT 1125

RESULT 9
US-10-229-546-1
; Sequence 1, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1622
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1254)
US-10-229-546-1

Alignment Scores:
Pred. No.:              6.51e-68        Length:         1622
Score:                  619.50          Matches:        130
Percent Similarity:     54.07%          Conservative:   76
Best Local Similarity:  34.12%          Mismatches:     128
Query Match:            32.42%          Indels:         47
DB:                     15              Gaps:           8

US-09-980-881A-2 (1-360) x US-10-229-546-1 (1-1622)

Qy          6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla--------- 22
              :::   :::::::::|||:::   |||                     |||
Db          1 ATGAGGCTCATCCTGCCTGTGGGTTTGATTGCTACCACTCTTGCAATTGCTCCTGTCCGC 60

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              |||          :::|||             |||:::      :::|||      ::::::
Db         61 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 120

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              ::::::|||      |||   |||:::      |||  |||                 :::
Db        121 AAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTA 180

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                        |||   |||   |||:::      :::         :::::::::
Db        181 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 240

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
                 |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db        241 GCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 300

Qy        103 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ::::::|||            :::   :::       |||   |||          |||
Db        301 GAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCA 351

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              :::|||::::::      :::|||   :::|||   |||  :::   :::::::::|||
Db        352 AAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCT 411

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
              :::|||::::::::::|||   |||||||||:::   |||      |||||||||||||||
Db        412 GAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTG 471

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              |||:::   |||::::::         :::   ||||||::::::|||||||||||||||
Db        472 AAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCAC 528

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------- 195
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        529 GCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACT 588

Qy        195 ------------------------------------------------------------ 195

Db        589 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 648

Qy        196 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204
                             |||:::                 ||||||||||||||||||
Db        649 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 708

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||      |||:::   ||||||||||||||||||||   ||||||   :::
Db        709 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 765

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
                    |||              ::::::      |||::::::   |||   |||
Db        766 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
                 ||||||||||||   |||   |||||||||   :::|||:::|||   ::::::|||
Db        826 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 885

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 284
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        886 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 945

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              ||||||                 :::|||||||:::|||:::   ||||||
Db        946 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1005

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
                  :::     |||          |||||||||   :::|||            |||
Db       1006 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1062

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              :::|||          |||          ||||||   ||||||||||||||||||:::::
Db       1063 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1122

Qy        345 Phe 345
              |||
Db       1123 TTT 1125

RESULT 10
US-10-200-344-9
; Sequence 9, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
```

```
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-200-344-9

Alignment Scores:
Pred. No.:                4.74e-67      Length:          1311
Score:                    611.50        Matches:         141
Percent Similarity:       50.36%        Conservative:    69
Best Local Similarity:    33.81%        Mismatches:      135
Query Match:              32.00%        Indels:          72
DB:                       14            Gaps:            11

US-09-980-881A-2 (1-360) x US-10-200-344-9 (1-1311)

Qy          1 MetLysLeuCys------SerLeuAlaValLeuValProIleValLeuPheCysGluGln 18
              :::  ||||||         |||  :::|||  |||
Db         43 CTGCCTCTTTGCTGGCTCTTTTTGAAGATTCTGCAACCG-----------GGGCACAGC 90

Qy         19 HisValPhe------AlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSer 36
              |||::::::        :::      :::|||:::      :::|||:::|||
Db         91 CACCTTTATAACAACCGCTATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAA 150

Qy         37 ArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnPro 56
                 :::       |||:::  ::::::        ::::::  ||||||||||||
Db        151 GAGGAAGCATATGCACTGAAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCC 210

Qy         57 ValThrAlaAspLeuIleValLys------LysLysGlnValHisPhePheValAsnAla 74
                 :::       :::  :::              ||||||         |||
Db        211 AGCAGTATCTCCTATGTATCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGT 270

Qy         75 SerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeu 94
              |||           :::  |||  |||      :::  |||         ||||||
Db        271 TCCCGAGCC------CTGTTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTC 324

Qy         95 LeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrValSerProArg 114
              :::  |||::::::      :::::::::      |||      |||         |||
Db        325 ATAGAAGATCTTCAGAAAACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGA 384

Qy        115 AlaSerAlaSer------TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrp 132
              |||  |||         ||||||  ||||||||||||  ||||||  :::|||
Db        385 AGATCCCTCTCTGGATATAATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGG 444

Qy        133 IleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSer 152
              :::      :::  :::  |||     ::::::              ||||||  |||
Db        445 ATGCATCATCTGAATAAAACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCA 504

Qy        153 PheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 172
              :::|||         |||::::::||||||:::  |||:::            |||
Db        505 TATGAGGGAAGATCTCTTTTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGA 561

Qy        173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeu 192
              |||:::||||||||||||||||||||||||||||||||  ||||||||||
Db        562 GCTGTTTGGATAGACTGTGGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAG 621

Qy        193 TrpPheIle---------------------------------------------------- 195
              ||||||:::
Db        622 TGGTTTGTAAAAGAAGCTCTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTG 681

Qy        195 ---------------------------------------------------------- 195

Db        682 AATCATCTATATTTCTATATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGG 741

Qy        196 GlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGly 215
                 ::::::|||   |||||||||   ||||||      :::         |||   |||

Db        742 ACCAATGATCGATTTTGGAGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGA 801

Qy        216 ThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSer 235
                 |||   |||   |||:::       |||   ||||||:::||||||||||||
Db        802 GTGGATGCCAATAGAAACTGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCAC 858

Qy        236 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaVal 255
                 |||   :::|||||||||||||   :::||||||||||||||||||||||||||||||
Db        859 CCTTGTGATGACACATACTGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTA 918

Qy        256 AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyr 275
                 |||:::|||||||||:::::          |||:::||||||:::|||   |||:::|||
Db        919 GCTAACTTCCTTCGAAAACACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATAT 978

Qy        276 SerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGlu 295
                 :::|||   :::::::::||||||||||||       :::         :::
Db        979 GCTCAGATGTTACTGTATCCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGT 1038

Qy        296 LeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArg 315
                 :::       |||   :::||||||   |||:::   ::::::          |||
Db       1039 GTGGAATCTGCAGCTTATAAAGCTGTGAATGCACTT---CAGTCAGTATACGGGGTACGA 1095

Qy        316 TyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrp 335
                 |||   :::|||       |||   |||||||||::::::   |||         ||||||
Db       1096 TACAGATATGGACCAGCCTCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGG 1155

Qy        336 IleTyrAspLeuGlyIleLysTyrSerPhe----------------------------- 345
                 |||       ||||||   |||:::|||
Db       1156 GCCTACAAAAATGGAATACCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTT 1215

Qy        346 ----------------------ThrSerAsnProProValGluLysLeu 354
                                     :::||||||||||||||:::||||||
Db       1216 GGATTTTTACTCCCAGAGATGCTCATCAAACCCACCTGTACAGAAACTA 1264

RESULT 11
US-10-274-639-33
; Sequence 33, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
```

```
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 33
;   LENGTH: 1993
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CB1
US-10-274-639-33

Alignment Scores:
Pred. No.:              9.62e-67        Length:         1993
Score:                  611.50          Matches:        141
Percent Similarity:     50.36%          Conservative:   69
Best Local Similarity:  33.81%          Mismatches:     135
Query Match:            32.00%          Indels:         72
DB:                     12              Gaps:           11

US-09-980-881A-2 (1-360) x US-10-274-639-33 (1-1993)

Qy          1 MetLysLeuCys------SerLeuAlaValLeuValProIleValLeuPheCysGluGln 18
              :::   ||||||         |||   :::|||   |||
Db        380 CTGCCTCTTTGCTGGCTCTTTTTGAAGATTCTGCAACCG-----------GGGCACAGC 427

Qy         19 HisValPhe------AlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSer 36
              |||::::::         :::         :::|||:::      :::|||:::|||
Db        428 CACCTTTATAACAACCGCTATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAA 487

Qy         37 ArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnPro 56
              :::         |||:::   ::::::       ::::::   ||||||||||||
Db        488 GAGGAAGCATATGCACTGAAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCC 547

Qy         57 ValThrAlaAspLeuIleValLys------LysLysGlnValHisPheValAsnAla 74
                 :::         :::   :::             ||||||         |||
Db        548 AGCAGTATCTCCTATGTATCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGT 607

Qy         75 SerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeu 94
              |||             :::   |||   |||       :::   |||         ||||||
Db        608 TCCCGAGCC------CTGTTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTC 661

Qy         95 LeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrValSerProArg 114
              :::   |||::::::      :::::::::       |||      |||         |||
Db        662 ATAGAAGATCTTCAGAAAACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGA 721

Qy        115 AlaSerAlaSer------TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrp 132
                 |||   |||         ||||||   ||||||||||||   ||||||   :::|||
Db        722 AGATCCCTCTCTGGATATAATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGG 781

Qy        133 IleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSer 152
              :::       :::   :::   |||       ::::::             ||||||   |||
Db        782 ATGCATCATCTGAATAAAACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCA 841

Qy        153 PheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsn 172
              :::|||         |||::::::||||:::   |||:::         |||
Db        842 TATGAGGGAAGATCTCTTTTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGA 898

Qy        173 AlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeu 192
              |||:::|||||||||||||||||||||||||||||||||||   ||||||||||||
Db        899 GCTGTTTGGATAGACTGTGGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAG 958

Qy        193 TrpPheIle---------------------------------------------------- 195
              ||||||:::
Db        959 TGGTTTGTAAAAGAAGCTCTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTG 1018

Qy        195 ------------------------------------------------------------ 195

Db       1019 AATCATCTATATTTCTATATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGG 1078

Qy        196 GlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGly 215
                 :::::|||   |||||||||   ||||||      :::         |||   |||

Db       1079 ACCAATGATCGATTTTGGAGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGA 1138

Qy        216 ThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSer 235
                 |||   |||   |||:::      |||   ||||||:::||||||||||||
Db       1139 GTGGATGCCAATAGAAACTGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCAC 1195

Qy        236 SerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaVal 255
                 |||   :::||||||||||||   :::||||||||||||||||||||||||||||||
Db       1196 CCTTGTGATGACACATACTGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTA 1255

Qy        256 AlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyr 275
              |||:::|||||||||:::::::         |||:::||||||:::|||   |||:::|||
Db       1256 GCTAACTTCCTTCGAAAACACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATAT 1315

Qy        276 SerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGlu 295
              :::|||   :::::::::||||||||||||       :::         :::
Db       1316 GCTCAGATGTTACTGTATCCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGT 1375

Qy        296 LeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArg 315
              :::         |||   :::||||||   |||:::   ::::::             |||
Db       1376 GTGGAATCTGCAGCTTATAAAGCTGTGAATGCACTT---CAGTCAGTATACGGGGTACGA 1432

Qy        316 TyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrp 335
              |||   :::|||      |||   ||||||||::::::   |||         ||||||
Db       1433 TACAGATATGGACCAGCCTCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGG 1492

Qy        336 IleTyrAspLeuGlyIleLysTyrSerPhe----------------------------- 345
                 |||       ||||||   |||:::|||
Db       1493 GCCTACAAAAATGGAATACCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTT 1552

Qy        346 ----------------------ThrSerAsnProProValGluLysLeu 354
                                     :::||||||||||||||:::||||||
Db       1553 GGATTTTTACTCCCAGAGATGCTCATCAAACCCACCTGTACAGAAACTA 1601

RESULT 12
US-09-960-352-14595
; Sequence 14595, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT:  Tao, Nengbing
;  APPLICANT:  Byatt, John C.
;  APPLICANT:  Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION:  MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14595
;   LENGTH: 416
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
US-09-960-352-14595

Alignment Scores:
Pred. No.:              2.76e-65        Length:         416
Score:                  591.00          Matches:        111
Percent Similarity:     89.13%          Conservative:   12
Best Local Similarity:  80.43%          Mismatches:     15
Query Match:            30.93%          Indels:         0
DB:                     10              Gaps:           0

US-09-980-881A-2 (1-360) x US-09-960-352-14595 (1-416)

Qy         51 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPhe 70
              ||||||||||||||||||||   |||:::   |||||||||      :::||||||||
Db          2 ATTGTTCTCTGGCAGCCAGTAGCAGCTGAATATATTGTGAAGGGATACGAAGTCCATTTT 61

Qy         71 PheValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIlePro 90
              ||||||||||||||||||||   ||||||||||||||||||||   |||   ||||||
Db         62 TTTGTGAATGCATCTGATGTAAGCAATGTGAAAGCCCATTTAAATGCGAGCAGAATTCCA 121
```

```
Qy          91 CysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThr 110
Db         122 TTCAGGGTCCTGGTGGAAAATGTGGAAGATCTTATCCGGCAGCAGACTTCCAATGACACC 181

Qy         111 ValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyr 130
Db         182 ATCAGCCCCCGGGCATCCTCCTCCTACTATGAACAGTATCACTCACTAAATGAGATCTAT 241

Qy         131 SerTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 150
Db         242 TCTTGGATAGAAGTTATGACTGAGCGGTATCCTGATATGGTTGAAAAAATCCACATTGGA 301

Qy         151 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAla 170
Db         302 TCCTCATACGAGAAGTACCCACTTTATGTTTTAAAGGTTCTAAAAAAGAACAAAGGGCC 361

Qy         171 LysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 188
Db         362 AAAAATGCCATGTGGATTGACTGTGGAATCCACGCCAGAGAGTGGATCTCTCCT 415

RESULT 13
US-10-200-344-13
; Sequence 13, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641Alel Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 2128
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-200-344-13

Alignment Scores:
Pred. No.:                1.19e-60      Length:          2128
Score:                    564.00        Matches:         142
Percent Similarity:       45.24%        Conservative:    67
Best Local Similarity:    30.74%        Mismatches:      133
Query Match:              29.51%        Indels:          120
DB:                       14            Gaps:            11

US-09-980-881A-2 (1-360) x US-10-200-344-13 (1-2128)

Qy           8 ValLeuValProIleValLeuPheCysGluGln---------HisValPheAla-PheGl 24
Db         400 ATAATAATCAGTTGCCTGGTATACTGCTCAATGATATTGCCACACGTAAGAACAAGCATA 459

Qy          24 nSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAs 44
Db         460 CAGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTGAAGAA 519

Qy          44 nLeuThr-ThrThrTyrGluIleVal-----------------LeuTrpGlnProValT 58
Db         520 AATATCCTATCAACTTAAGGTTGGTTCCTGCACTACAGGTGGACCTGTGGCAGCCCAGCA 579

Qy          58 hrAlaAspLeuIleValLys------LysLysGlnValHisPhePheValAsnAlaSerA 76
Db         580 GTATCTCCTATGTATCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCC 639

Qy          76 spValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuA 96
Db         640 GAGCC------CTGTTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAG 693

Qy          96 laAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaS 116
Db         694 AAGATCTTCAGAAAACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGAT 753

Qy         116 erAlaSer------TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleG 134
Db         754 CCCTCTCTGGATATAATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGC 813

Qy         134 luPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheG 154
Db         814 ATCATCTGAATAAAACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATG 873

Qy         154 luLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaI 174
Db         874 AGGGAAGATCTCTTTTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTG 930

Qy         174 leTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpP 194
Db         931 TTTGGATAGACTGTGGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGT 990

Qy         194 heIle--------------------GlyHis-Asn---------------------- 198
Db         991 TTGTAAAAGAAGTCCTAGAAAACACAGCTCACAAATGTCAAGAATGTACTAAATTTACAA 1050

Qy         198 ---------------------------------------------------------- 198
Db        1051 AATATCTCTGCCACTACCAAAACCACAAAAGTATGCTTAATCTTGTAAGTATTGAGTAAT 1110

Qy         198 ---------------------------------------------------------- 198
Db        1111 AAAATTTTCTAAACATTCCTAAAAACTCTTCTAACATATAAGAGTGACCCAGCCATGAGA 1170

Qy         198 ---------------------------------------------------------- 198
Db        1171 AAAATGTTGAATCATCTATATTTCTATATCATGCCTGTGTTTAACGTCGATGGATACCAT 1230

Qy         199 ------------------ArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHis 212
Db        1231 TTTAGTTGGACCAATGATCGATTTTGGAGAAAAACAAGGTCAAGGAACTCAAGGTTTCGC 1290

Qy         213 CysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGluGlyAla 232
Db        1291 TGCCGTGGAGTGGATGCCAATAGAAACTGGAAAGTGAAG---TGGTGTGATGAAGGAGCT 1347

Qy         233 SerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluVal 252
Db        1348 TCTATGCACCCTTGTGATGACACATACTGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTG 1407

Qy         253 LysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMet 272
Db        1408 AAGGCTGTAGCTAACTTCCTTCGAAAACACAGAAAGCACATTAGGGCTTATCTCTCCTTT 1467

Qy         273 HisSerTyrSerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAsp 292
Db        1468 CATGCATATGCTCAGATGTTACTGTATCCCTATTCTTACAAATATGCAACAATTCCCAAT 1527

Qy         293 HisGluGluLeuSerLeuValAlaSerGluAlaValArgAlaIleAspLysThrSerLys 312
Db        1528 TTTAGATGTGTGGAATCTGCAGCTTATAAAGCTGTGAATGCACTT---CAGTCAGTATAC 1584

Qy         313 AsnThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGly 332
Db        1585 GGGGTACGATACAGATATGGACCAGCCTCCACAACGTTGTATGTGAGCTCTGGTAGCTCA 1644

Qy         333 AspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe---------------------- 345
Db        1645 ATGGATTGGGCCTACAAAAATGGAATACCTTATGCATTTGCTTTCGAACTACGTGACACT 1704

Qy         346 -----------------------------ThrSerAsnProProValGluLysLeu 354
```

```
                                                    :::|||||||||||||||:::||||||
Db       1705 GGATATTTTGGATTTTTACTCCCAGAGATGCTCATCAAACCCACCTGTACAGAAACTA 1762

RESULT 14
US-09-888-615-2
; Sequence 2, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1125
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-888-615-2

Alignment Scores:
Pred. No.:              5.43e-61        Length:         1125
Score:                  563.00          Matches:        109
Percent Similarity:     56.74%          Conservative:   51
Best Local Similarity:  38.65%          Mismatches:     82
Query Match:            29.46%          Indels:         40
DB:                     9               Gaps:           5

US-09-980-881A-2 (1-360) x US-09-888-615-2 (1-1125)

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr---TyrGlu 121
              :::|||:::|||   :::   :::|||||||||   :::      :::|||   |||
Db         85 AGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAAC 144

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              ||||||   :::   |||||||||   |||:::      |||:::|||::::::
Db        145 ATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAG 204

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
              :::::::::|||:::      :::|||   ::::::|||   :::|||:::|||   |||
Db        205 GAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTG 264

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              |||:::|||                  |||      ||||||:::|||||||||||||||
Db        265 AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCAC 324

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 195
              |||||||||||||||:::||||||||||||   ||||||:::
Db        325 GCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAAC 384

Qy        195 ------------------------------------------------------------ 195

Db        385 CATAAAGACAACTCAAGTATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCA 444

Qy        196 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204
                           |||:::                 :::|||:::|||||||||:::
Db        445 GTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC 504

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||   :::   |||       |||   ||||||||||||||   ||||||   :::
Db        505 CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC---AAT 561

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
              ||||||      |||||||||   :::      :::|||:::||||||
Db        562 GCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACA 621

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
              |||   ||||||||||||   ||||||||||||||||||:::           :::
Db        622 GGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGAT 681

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284
              |||      :::::::::||||||||||||   |||   |||:::   ||||||
Db        682 GATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGC 741

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db        742 TACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCA 801

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
              |||:::   |||          |||   |||      |||      ::::::
Db        802 AATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATT 858

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              ||||||   :::   |||         ||||||      |||:::||||||   :::|||
Db        859 TTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCA 918

Qy        345 PheThr 346
              :::|||
Db        919 TATACG 924

RESULT 15
US-10-176-306-75
; Sequence 75, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
```

```
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 75
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-176-306-75

Alignment Scores:
Pred. No.:              7.23e-61        Length:         1332
Score:                  563.00          Matches:        109
Percent Similarity:     56.74%          Conservative:   51
Best Local Similarity:  38.65%          Mismatches:     82
Query Match:            29.46%          Indels:         40
DB:                     16              Gaps:           5

US-09-980-881A-2 (1-360) x US-10-176-306-75 (1-1332)

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr---TyrGlu 121
              :::|||:::|||   :::   :::|||||||||   :::      :::|||   |||
Db        475 AGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAAC 534

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
                ||||||   :::   |||||||||   |||:::      |||:::|||::::::
Db        535 ATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAG 594

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
              :::::::::|||:::      :::|||   ::::::|||   :::|||:::|||   |||
Db        595 GAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATATATTATCTG 654

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              |||:::|||                |||      ||||||:::|||||||||||||||
Db        655 AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCAC 714

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----------------- 195
              |||||||||||||||:::|||||||||||||   ||||||:::
Db        715 GCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAAC 774

Qy        195 ------------------------------------------------------------ 195

Db        775 CATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCA 834

Qy        196 ---------------GlyHis-----------------AsnArgMetTrpArgLysAsn 204
                            |||:::                 :::|||:::|||||||||:::
Db        835 GTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC 894

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||   :::   |||      |||   ||||||||||||||   ||||||   :::
Db        895 CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC---AAT 951

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
                  ||||||      |||||||||   :::         :::|||:::||||||
Db        952 GCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACA 1011

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
                |||   ||||||||||||   ||||||||||||||||||:::             :::
Db       1012 GGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGAT 1071

Qy        265 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 284
                |||      :::::::::||||||||||||   |||   |||:::   ||||||
Db       1072 GATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGC 1131

Qy        285 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db       1132 TACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCA 1191

Qy        305 ArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 324
                |||:::   |||            |||   |||      |||      ::::::
Db       1192 AATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATT 1248

Qy        325 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 344
              ||||||   :::   |||         ||||||      |||:::||||||   :::|||
Db       1249 TTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCA 1308

Qy        345 PheThr 346
                :::|||
Db       1309 TATACG 1314


Search completed: January 2, 2004, 10:04:48
Job time : 908.574 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:47:17 ; Search time 3188.89 Seconds
(without alignments)
2743.777 Million cell updates/sec

Title: US-09-980-881A-2
Perfect score: 1911
Sequence: 1 MKLCSLAVLVPIVLFCEQHV.....IKYSFTSNPPVEKLLPLSLK 360

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135916_26873/app_query.fasta_1.1628
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_5167_@runat_31122003_135916_26873 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1534.5	80.3	1446	11	AK004045	AK004045 Mus muscu
	2	1351.5	70.7	971	9	AL536069	AL536069 AL536069
	3	1034	54.1	788	10	BG618629	BG618629 602646064
	4	1016	53.2	702	2	HSM089979	Bx496754 Homo sapi
	5	1015.5	53.1	853	12	BI144644	BI144644 602909932
	6	996.5	52.1	738	10	BG616456	BG616456 602614618
	7	989	51.8	1004	14	BY704293	BY704293 BY704293
	8	983	51.4	747	10	BG566051	BG566051 602582652
	9	975	51.0	725	10	BG618239	BG618239 602645967
	10	965	50.5	852	10	BG569281	BG569281 602588573
	11	957.5	50.1	1022	12	BI332659	BI332659 602984408
	12	949	49.7	735	9	AV646979	AV646979 AV646979
	13	927.5	48.5	750	10	BG618813	BG618813 602646186
	14	921	48.2	541	14	CB157619	CB157619 K-EST0216
	15	902	47.2	747	10	BG565100	BG565100 602583720
	16	898	47.0	691	10	BG568240	BG568240 602587149
	17	884.5	46.3	724	9	AI048167	AI048167 ud71b12.y
	18	868	45.4	583	9	AV693037	AV693037 AV693037
	19	857	44.8	602	9	AV692032	AV692032 AV692032
	20	849	44.4	601	9	AV658390	AV658390 AV658390
	21	848	44.4	623	9	AV694058	AV694058 AV694058
	22	838	43.9	621	9	AV698425	AV698425 AV698425
	23	837.5	43.8	749	10	BG617911	BG617911 602645566
	24	826	43.2	485	13	BX091627	BX091627 BX091627
	25	818.5	42.8	688	9	AV651709	AV651709 AV651709
	26	815.5	42.7	806	10	BG567264	BG567264 602589745
	27	814.5	42.6	577	14	CB154355	CB154355 K-EST0212
	28	806.5	42.2	661	9	AI529872	AI529872 ui83d11.y
	29	806	42.2	580	9	AI182113	AI182113 ud73f10.y
	30	806	42.2	589	9	AA968307	AA968307 uc71h07.y
	31	803.5	42.0	603	10	AW950700	AW950700 EST362770
c	32	799.5	41.8	973	9	AL536068	AL536068 AL536068
	33	782	40.9	510	14	CB156851	CB156851 K-EST0215
	34	773.5	40.5	537	14	N98450	N98450 za28c09.r1
	35	767	40.1	958	9	AI322676	AI322676 mj24d06.y
	36	766	40.1	511	9	AI574117	AI574117 uj68d09.y
	37	765.5	40.1	787	14	CB593837	CB593837 AGENCOURT
	38	759	39.7	511	9	AI528146	AI528146 ui94a04.y
	39	742	38.8	462	9	AV681868	AV681868 AV681868
c	40	741.5	38.8	843	10	BG216526	BG216526 RST36106
	41	731	38.3	455	14	W88434	W88434 zh69f04.r1
	42	728.5	38.1	623	9	AV658178	AV658178 AV658178
	43	719	37.6	434	14	CB164014	CB164014 K-EST0225
	44	718	37.6	615	9	AV658128	AV658128 AV658128
	45	711.5	37.2	558	4	BX528015	Bx528015 RZPD Mus

ALIGNMENTS

RESULT 1
AK004045
LOCUS AK004045 1446 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1110032P04 product:carboxypeptidase B2 (plasma), full insert sequence.
ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1446)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1446
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1110032P04"
/db_xref="MGI:1900711"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
CDS 42. .1310
/note="unnamed protein product; carboxypeptidase B2 (plasma) (MGD|MGI:1891837, GB|NM_019775, evidence: BLASTN, 99%, match=1419)
putative"
/codon_start=1
/protein_id="BAB23141.1"
/db_xref="GI:12835068"
/db_xref="MGI:1891837"
/translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT 421 a 329 c 311 g 385 t
ORIGIN

Alignment Scores:

Pred. No.:	9.3e-165	Length:	1446
Score:	1534.50	Matches:	302
Percent Similarity:	78.80%	Conservative:	25
Best Local Similarity:	72.77%	Mismatches:	32
Query Match:	80.30%	Indels:	56
DB:	11	Gaps:	3

US-09-980-881A-2 (1-360) x AK004045 (1-1446)

```
Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||     |||   :::||||||   |||:::|||:::   |||||||||
Db         42 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTAT---GAGCAGCATGGC 98

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||:::||||||||||||||||||||||||||||||
Db         99 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 158

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||||:::
Db        159 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 218
```

```
Qy          61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
               |||   |||||||||:::|||||||||||||||||||||||||||||||||:::|||
Db         219 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 278

Qy          81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
               ||||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
Db         279 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 338

Qy         101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
               ||||||:::||||||      ||||||||||||||||||||||||||||||||||||||
Db         339 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 398

Qy         121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
               ||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db         399 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 458

Qy         141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
               ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db         459 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 518

Qy         161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
               ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db         519 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 578

Qy         181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
               ||||||||||||||||||||||||||||||||||||||||||||||:::
Db         579 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 638

Qy         197 ------------------------------------------------------------ 197

Db         639 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 698

Qy         198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db         699 CCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAG 758

Qy         204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
               |||||||||   :::   ||||||   |||:::||||||||||||||   ||||||
Db         759 AACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 818

Qy         224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
               ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db         819 TCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGA 878

Qy         244 LeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIle 263
               ||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db         879 CTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATC 938

Qy         264 AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyr 283
               :::   |||||||||||||||||||||||||||||||||||   |||:::|||||||||
Db         939 GACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTAT 998

Qy         284 SerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAla 303
               ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         999 TCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCA 1058

Qy         304 ValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGlu 323
               ||||||||||:::      :::|||||||||||||||||||||||||   |||||||||
Db        1059 GTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAA 1118

Qy         324 ThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyr 343
               :::|||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        1119 AGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATAT 1178

Qy         344 SerPhe-------------------------------------------------------Th 346
               ||||||                                                       ||
Db        1179 TCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATAC 1238

Qy         346 rSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 360
               ||||||||||   |||:::||||||   |||   |||||||||
Db        1239 ATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1281

RESULT 2
```

```
AL536069
LOCUS       AL536069                971 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION   AL536069
VERSION     AL536069.2  GI:31260939
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12799562.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6944.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF022AC11QP1&cluster=6944.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF022AC11QP1.
FEATURES             Location/Qualifiers
     source          1. .971
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF022YE21"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
BASE COUNT      282 a    201 c    211 g    272 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:              4.91e-144       Length:         971
Score:                  1351.50         Matches:        259
Percent Similarity:     86.33%          Conservative:   0
Best Local Similarity:  86.33%          Mismatches:     4
Query Match:            70.72%          Indels:         37
DB:                     9               Gaps:           1

US-09-980-881A-2 (1-360) x AL536069 (1-971)

Qy           2 LysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPhe 21
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          68 AAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTC 127

Qy          22 AlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnVal 41
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         128 GCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTT 187

Qy          42 LeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeu 61
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         188 CTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTT 247

Qy          62 IleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLys 81
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         248 ATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAA 307

Qy          82 AlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeu 101
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        308 GCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTT 367

Qy        102 IleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        368 ATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAA 427

Qy        122 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 141
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        428 CAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCCT 487

Qy        142 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 161
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        488 GATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTA 547

Qy        162 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        548 AAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 607

Qy        182 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        608 GCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTC 667

Qy        197 ------------------------------------------------------------ 197

Db        668 TATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCG 727

Qy        198 ----------------------------------------AsnArgMetTrpArgLysAsn 204
                                                     ||||||||||||||||||||
Db        728 GTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAAC 787

Qy        205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSer 224
              ||||||||||||||||||||||||||||||||||||||||||||||    ||||||   |||
Db        788 CGTTCTTTCTATGCGAACAATCATTGCATTGGAACAGACCTGAATAGGAACTTTGCTTCC 847

Qy        225 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 244
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        848 AAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTT 907

Qy        245 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        908 TATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAWATMAAC 967

RESULT 3
BG618629
LOCUS       BG618629                 788 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
            mRNA sequence.
ACCESSION   BG618629
VERSION     BG618629.1  GI:13670000
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: o  column: 15
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767974"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      222 a    157 c    169 g    240 t
ORIGIN

Alignment Scores:
Pred. No.:              9.42e-108      Length:          788
Score:                  1034.00        Matches:         201
Percent Similarity:     83.82%         Conservative:    1
Best Local Similarity:  83.40%         Mismatches:      1
Query Match:            54.11%         Indels:          38
DB:                     10             Gaps:            1

US-09-980-881A-2 (1-360) x BG618629 (1-788)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         66 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 125

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        126 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 185

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        186 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 245

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        246 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 305

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        306 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 365

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        366 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 425

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        426 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 485

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        486 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 545

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        546 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 605

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        606 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAA 665

Qy        197 ------------------------------------------------------------ 197

Db        666 TTCTATGGGATAATAGGGCAATATTACCAATCTCCTGAGGCTTGTGGATTTCTATGTTAT 725

Qy        198 ----------------------------------------------AsnArgMetTrpArg 202
                                                           ||||||:::||||||
```

```
Db        726 GCCAGTGGTTAATGTTGGATGGGTTATGACTACTCATGGAAAAGGAATCGATTGTGGAGA 785

Qy        203 Lys 203
                |||
Db        786 AAG 788

RESULT 4
HSM089979
ID   HSM089979  standard; RNA; EST; 702 BP.
XX
AC   BX496754;
XX
SV   BX496754.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp779I1529_r1  (from clone DKFZp779I1529)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-702
RA   Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by EMBL (European Molecular Biology Laboratories,
CC   Heidelberg/Germany) within the cDNA sequencing consortium of
CC   the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp779I1529) is available at the RZPD in Berlin.
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .702
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp779I1529"
FT                   /clone_lib="779 (synonym: hncc1). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="fetal"
FT                   /tissue_type="liver"
XX
SQ   Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;

Alignment Scores:
Pred. No.:              9.12e-106      Length:         702
Score:                  1016.00        Matches:        196
Percent Similarity:     98.99%         Conservative:   0
Best Local Similarity:  98.99%         Mismatches:     1
Query Match:            53.17%         Indels:         1
DB:                     2              Gaps:           0

US-09-980-881A-2 (1-360) x HSM089979 (1-702)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         59 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 118

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        119 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 178

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        179 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 238

Qy         61 LeuIleValLys-LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVa 80
              |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db        239 CTTATTGTGAAGAAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGT 298

Qy         80 lLysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAs 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        299 GAAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGA 358

Qy        100 pLeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTy 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 TCTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTA 418

Qy        120 rGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHi 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 TGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCA 478

Qy        140 sProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVa 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 TCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGT 538

Qy        160 lLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIl 180
              |||||||||||||||||||||||    ||||||||||||||||||||||||||||||||
Db        539 TTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAAT 598

Qy        180 eHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 CCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 650

RESULT 5
BI144644
LOCUS       BI144644                 853 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602909932F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5051358 5',
            mRNA sequence.
ACCESSION   BI144644
VERSION     BI144644.1  GI:14604645
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11139  row: o  column: 07
            High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5051358"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
```

```
                Average insert size 1.9 kb. Constructed by Life
                Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      243 a    199 c    194 g    217 t
ORIGIN

Alignment Scores:
Pred. No.:              1.41e-105       Length:         853
Score:                  1015.50         Matches:        202
Percent Similarity:     76.66%          Conservative:   18
Best Local Similarity:  70.38%          Mismatches:     26
Query Match:            53.14%          Indels:         41
DB:                     12              Gaps:           3

US-09-980-881A-2 (1-360) x BI144644 (1-853)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              |||      |||   :::||||||   |||:::   :::   |||||||||   ||||||
Db          2 CTTCATGGCCTTGGAATCCTGGTAGCCATCATCCGCTAT---GAGCAGCATGGCTTCGCC 58

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||||:::|||
Db         59 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 118

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              ||||||||||||||||||||||:::|||||||||||||||||||||||:::   |||
Db        119 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 178

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
                 ||||||||:::|||||||||||||||||||||||||||||||:::||||||||||
Db        179 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 238

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
              ||||||||||||||||   ||||||   :::||||||:::   :::|||||||||||||||
Db        239 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 298

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
              :::||||||      ||||||||||||||||||||||||||||||||||||||||||||
Db        299 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 358

Qy        123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||:::||||||||
Db        359 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 418

Qy        143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
              ||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        419 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 478

Qy        163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        479 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 538

Qy        183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 197
              |||||||||||||||||||||||||||||||||||||||||:::
Db        539 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 598

Qy        197 ------------------------------------------------------------ 197

Db        599 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 658

Qy        198 ----------------------------------AsnArgMetTrpArgLysAsnArg 205
                                                 ||||||||||||||||||||||
Db        659 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAACCGT 718

Qy        206 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLys 225
                          |||:::|||   |||||||||   ||||||   ||||||
Db        719 ---CTGTCACAAGAACAACCGTGCGTTGGGCCAGACCTGAACAGGAACTTCGCTTCCAAA 775

Qy        226 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGlu-ThrTyrCys-GlyLeuT 245
              ||||||||||||:::|||||||||||||||||||||||||| |||||||| |||||||
Db        776 CACTGGTGTGAGAAAGGTGCGTCCAGTTCCTCCTGCTCTGAAAACCTACTGTGGGACTTT 835

Qy        245 yrProGluSerGluPro 250
              ||||||||||||||||
```

```
Db        836 ATCCTGAGTCTGAGCCG 852

RESULT 6
BG616456
LOCUS       BG616456                738 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
            mRNA sequence.
ACCESSION   BG616456
VERSION     BG616456.1  GI:13667827
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1595  row: c  column: 05
            High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4733500"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      209 a    161 c    153 g    215 t
ORIGIN

Alignment Scores:
Pred. No.:              1.7e-103        Length:         738
Score:                  996.50          Matches:        205
Percent Similarity:     72.39%          Conservative:   10
Best Local Similarity:  69.02%          Mismatches:     27
Query Match:            52.15%          Indels:         55
DB:                     10              Gaps:           3

US-09-980-881A-2 (1-360) x BG616456 (1-738)

Qy         10 ValProIleValLeuPheCysGluGlnHisValPheAlaPheGlnSerGlyGlnValLeu 29
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTTCAGAGTGGCCAAGTTCTA 60

Qy         30 AlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsnLeuThrThrThrTyr 49
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTTACTACAACATAT 120

Qy         50 GluIleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGlnValHis 69
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATTGTGAAGAAAAAACAAGTCCAT 180

Qy         70 PhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIle 89
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATT 240

Qy         90 ProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAsp 109
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAGATTTCCAACGAC 300

Qy        110 ThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIle 129
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATC 360

Qy        130 TyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIle 149
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATT 420

Qy        150 GlySerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThr 169
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCA 480

Qy        170 AlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAla 189
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCT 540

Qy        190 PheCysLeuTrpPheIleGlyHis--AsnArgMetTrpArgLysAsnArgSerPheTyrA 209
                ||||||||||||||||||||||  |||   :::         ||||||:::   |||
Db        541 TTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACC 600

Qy        209 laAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysG 229
              :::           |||                                       |||
Db        601 AATCTCCTGAGGCTTGTGGA----------------------------TTTCTATGTT 630

Qy        229 luGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerG 249
              |||   |||                          |||||||||
Db        631 ATGCCGGTGGTTCA----------------------TGTGGACTG------------ 653

Qy        249 luProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaT 269

Db        653 ------------------------------------------------------------ 653

Qy        269 yrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyrThrArgSerL 289
                                          |||      ||||||:::|||:::::::
Db        654 ------------------------------GTTATGGACTACTCATGGACCAAGAATC 681

Qy        289 ysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 304
              :::::         :::|||      |||         |||   :::
Db        682 GCACTGTTGGACAAAAGGAACCCGGTCTTTACTAATGCGAAACAATT 728
```

RESULT 7
BY704293

```
LOCUS       BY704293               1004 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION  BY704293 RIKEN full-length enriched, 18-day embryo whole body Mus
            musculus cDNA clone 1110032P04 5', mRNA sequence.
ACCESSION   BY704293
VERSION     BY704293.1  GI:27115417
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
    TITLE   Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1004
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1110032P04"
                     /tissue_type="whole body"
                     /dev_stage="18-day embryo"
                     /clone_lib="RIKEN full-length enriched, 18-day embryo
                     whole body"
BASE COUNT      286 a    228 c    228 g    262 t
ORIGIN
```

Alignment Scores:

Pred. No.:	2.01e-102	Length:	1004
Score:	989.00	Matches:	199
Percent Similarity:	75.69%	Conservative:	19
Best Local Similarity:	69.10%	Mismatches:	31
Query Match:	51.75%	Indels:	42

```
DB:                 14          Gaps:          3

US-09-980-881A-2 (1-360) x BY704293 (1-1004)

Qy           1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
               |||||||||      |||   :::||||||    |||:::|||:::   |||||||||
Db          42 ATGAAGCTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTAT---GAGCAGCATGGC 98

Qy          21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
               ||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||
Db          99 TTCGCCTTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAA 158

Qy          41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
               :::|||||||||||||||||||||||||:::|||||||||||||||||||||||||:::
Db         159 CTACTTCAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAA 218

Qy          61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
                  |||   |||||||||:::||||||||||||||||||||||||||||||||:::|||
Db         219 TTCATCGAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTC 278

Qy          81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
               ||||||||||||||||||||   ||||||   :::||||||:::   :::|||||||||
Db         279 AAAGCGCATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGAC 338

Qy         101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
               ||||||:::||||||      ||||||||||||||||||||||||||||||||||||||
Db         339 CTAATTGAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTAT 398

Qy         121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
               ||||||||||||||||||||||||||||||||||||||||   |||||||||:::|||
Db         399 GAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCAT 458

Qy         141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
               ||||||||||||   ||||||:::||||||||||||||||||||||||||||||||||||
Db         459 CCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTT 518

Qy         161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
               ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db         519 TTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATC 578

Qy         181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------- 197
               ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db         579 CATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAA 638

Qy         197 ------------------------------------------------------------ 197

Db         639 TTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATG 698

Qy         198 ----------------------------------------AsnArgMetTrpArgLys 203
                                                       ||||||||||||||||
Db         699 CGCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGTATGTGGAGGAAG 758

Qy         204 AsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnSerAsnPheVal 223
                  ||||||   :::   ||||||   |||:::|||||||||||||   ||||||
Db         759 AAGCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCT 818

Qy         224 SerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGly 243
               ||||||||||||||||:::                  ||||||   ||||||||||||
Db         819 TCCAAACACTGGTGTGAGAAG---GTGCGTAAGTTCCTCTGCTCTGA-ACCTACTGTGGA 874

Qy         244 LeuTyrProGluSerGluProGlu 251
                  |||       ||||||||||||
Db         875 TT-TATTCTAG-TCTGAGCCAGAG 896

RESULT 8
BG566051
LOCUS       BG566051                747 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602582652F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
            mRNA sequence.
ACCESSION   BG566051
VERSION     BG566051.1  GI:13573704
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1549  row: g  column: 13
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4710564"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    157 c    159 g    219 t
ORIGIN

Alignment Scores:
Pred. No.:              6.14e-102       Length:         747
Score:                  983.00          Matches:        199
Percent Similarity:     92.66%          Conservative:   3
Best Local Similarity:  91.28%          Mismatches:     13
Query Match:            51.44%          Indels:         4
DB:                     10              Gaps:           0

US-09-980-881A-2 (1-360) x BG566051 (1-747)

Qy           1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
               ||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db          21 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGT-CTCTTCTGTGAGCAGCATGTC 79

Qy          21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
               ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db          80 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTATAACCTCTAGGCAAGTTCAA 139

Qy          41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         140 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 199

Qy          61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         200 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 259

Qy          81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         260 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 319

Qy         101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         320 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 379
```

```
Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        380 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 439

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        440 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 499

Qy        161 -LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIl 180
               ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||
Db        500 TCTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAAT 559

Qy        180 eHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--AsnArg 199
              ||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db        560 CCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCA 619

Qy        200 MetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGly 215
              :::     ||||||:::   |||   :::       |||
Db        620 ATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGA 667
```

RESULT 9
BG618239
LOCUS BG618239 725 bp mRNA linear EST 18-APR-2001
DEFINITION 602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5', mRNA sequence.
ACCESSION BG618239
VERSION BG618239.1 GI:13669610
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 725)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1629 row: f column: 17
High quality sequence stop: 723.

FEATURES Location/Qualifiers
source 1. .725
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4767376"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_76"
/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 203 a 153 c 149 g 220 t
ORIGIN

Alignment Scores:

Pred. No.:	4.85e-101	Length:	725
Score:	975.00	Matches:	194
Percent Similarity:	98.48%	Conservative:	0
Best Local Similarity:	98.48%	Mismatches:	3
Query Match:	51.02%	Indels:	2
DB:	10	Gaps:	0

US-09-980-881A-2 (1-360) x BG618239 (1-725)

```
Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 124

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 184

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        185 GTTCTACAGAATCTTACTACAACATATGAGAT-GTTCTCTGGCAGCCGGTAACAGCTGAC 243

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        244 CTTAT-GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 302

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGAT 362

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTAT 422

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 482

Qy        141 ProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        483 CCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTT 542

Qy        161 LeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIle 180
              ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        543 TTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATC 602

Qy        181 HisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 CATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 653
```

RESULT 10
BG569281
LOCUS BG569281 852 bp mRNA linear EST 10-APR-2001
DEFINITION 602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5', mRNA sequence.
ACCESSION BG569281
VERSION BG569281.1 GI:13576934
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1580 row: b column: 19
High quality sequence stop: 751.

FEATURES Location/Qualifiers

```
     source          1. .852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4722354"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      251 a    177 c    187 g    237 t
ORIGIN

Alignment Scores:
Pred. No.:              8.78e-100       Length:         852
Score:                  965.00          Matches:        194
Percent Similarity:     75.38%          Conservative:   2
Best Local Similarity:  74.62%          Mismatches:     7
Query Match:            50.50%          Indels:         58
DB:                     10              Gaps:           2

US-09-980-881A-2 (1-360) x BG569281 (1-852)

Qy        157 ProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIle 176
              ||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db          2 CCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATT 61

Qy        177 AspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGly 196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGC 121

Qy        197 His---------------------------------------------------------- 197
              |||
Db        122 CATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGAT 181

Qy        198 ----------------------------------------------------------AsnArg 199
                                                                         ||||||
Db        182 TTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGA 241

Qy        200 MetTrp-ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAs 219
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ATGTGGACGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAA 301

Qy        219 nSerAsnPheValSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGl 239
              |   ||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGA 361

Qy        239 uThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLe 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTT 421

Qy        259 uArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIl 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATAT 481

Qy        279 eValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVa 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGT 541

Qy        299 lAlaSerGluAlaValArgAlaIleAspLysThrSerLysAsnThrArgTyrThrHisGl 319
              |||||||||||||||||||||||:::|||   ||||||||||||||||||||||||||||
Db        542 AGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGC 601

Qy        319 yHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr-AspL 339
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||    |
Db        602 GCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGCATT 661

Qy        339 euGlyIleLysTyrSerPhe---------------------------------------- 345
              ||||||||:::|||||||||
Db        662 TGGGCATCAGATATTCGTTTACAATTGAACCTTCGAGATACGGGCACATAACGGATTCTT 721

Qy        346 --------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 359
                           ||||||   ||||||||||||||||||||||||||||||||
Db        722 GCTGCCGGAGCGTTACATCAAA-CCACCTGTAGAGAAGCTTTGCCGCTGTCTCTA 776

RESULT 11
BI332659
LOCUS       BI332659                1022 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  602984408F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5137369 5',
            mRNA sequence.
ACCESSION   BI332659
VERSION     BI332659.1  GI:15017316
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1022)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11337  row: o  column: 02
            High quality sequence stop: 841.
FEATURES             Location/Qualifiers
     source          1. .1022
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5137369"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      303 a    233 c    225 g    261 t
ORIGIN

Alignment Scores:
Pred. No.:              8.49e-99        Length:         1022
Score:                  957.50          Matches:        209
Percent Similarity:     72.81%          Conservative:   24
Best Local Similarity:  65.31%          Mismatches:     44
Query Match:            50.10%          Indels:         46
DB:                     12              Gaps:           3

US-09-980-881A-2 (1-360) x BI332659 (1-1022)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              |||      |||   :::||||||   |||:::|||:::   |||||||||   ||||||
Db          3 CTTCATGGCCTTGGAATCCTGGTAGCCATCATCCTCTAT---GAGCAGCATGGCTTCGCC 59

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||:::|||
Db         60 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 119

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              |||||||||||||||||||||:::|||||||||||||||||||||||:::   |||
Db        120 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 179
```

```
Qy          63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
Db         180 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 239

Qy          83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
Db         240 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 299

Qy         103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
Db         300 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 359

Qy         123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
Db         360 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 419

Qy         143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
Db         420 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 479

Qy         163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 182
Db         480 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 539

Qy         183 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 197
Db         540 AGAGAATGGATTTCACCTGCTTTCTGTTG-TGGTTCATAGGCTACGTGACACAATTCCAT 598

Qy         197 ------------------------------------------------------------ 197

Db         599 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGAATTTCTACATCATGCCCGT 658

Qy         198 ----------------------------------------AsnArgMetTrpArgLysAsn 204
Db         659 GATGGAACGTGGATGGCTATGACTACACGTGGACAAAAGAATCGAATGTTGCAGGAAGAA 718

Qy         205 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThr---AspLeuAsnSerAsnPheVal 223
Db         719 CCGTCTGCTCACCAGAACAACCGCTGCTGTGGGCAACATGACCTGAACAGGAACTTCGCT 778

Qy         224 SerLysHisTrpCysGluGluGly-AlaSerSerSerSerCysSerGluThrTyrCysGl 243
Db         779 TCCAAACACTGGTGTGAGAAAGGTTGCGTCAAGTTTCCTCCTGTCTGAAAC-TACTGTGG 837

Qy         243 yLeuTyr-ProGluSer-GluProGluValLysAlaValAlaSerPheLeuArgArgAsn 262
Db         838 ACTTTATTCCCGAGTCTAGAGCCAGAAGTGAAGGCCAGGGGTGACTTTTTGAGAAGAATA 897

Qy         263 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 280
Db         898 TCTGACCCCCTTAAAGT-TACTTCAAGATGGCTCATAATTCCACAAATATGGTT 950
```

```
RESULT 12
AV646979
LOCUS       AV646979                735 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.
ACCESSION   AV646979
VERSION     AV646979.1  GI:9867993
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCATD06"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      200 a    158 c    172 g    205 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                4.77e-98        Length:          735
Score:                    949.00          Matches:         187
Percent Similarity:       96.89%          Conservative:    0
Best Local Similarity:    96.89%          Mismatches:      5
Query Match:              49.66%          Indels:          1
DB:                       9               Gaps:            0

US-09-980-881A-2 (1-360) x AV646979 (1-735)

Qy           6 LeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAlaPheGlnSer 25
Db           7 CTTGCGGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCGTTTCAGAGT 66

Qy          26 GlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsnLeu 45
Db          67 GGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTACAGAATCTT 126

Qy          46 ThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIleValLysLys 65
Db         127 ACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGGTGACCTTATTGTGAAGAAA 186

Qy          66 LysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsn 85
Db         187 AAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAAT 246

Qy          86 ValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGln 105
Db         247 GTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATTCAACAGCAG 306

Qy         106 IleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSer 125
Db         307 ATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGGACTATGAACAGTATCACTCA 366

Qy         126 LeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAspMetLeuThr 145
Db         367 CTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACA 426

Qy         146 LysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys-ValSerGl 165
Db         427 AAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAGGGTGTCTGG 486

Qy         165 yLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTr 185
Db         487 AAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATG 546

Qy         185 pIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
```

```
Db        547 GATCTCTCCTGCTTTCTGGTTGGGGGTCATAGGCCAT 583

RESULT 13
BG618813
LOCUS       BG618813                 750 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646186F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767885 5',
            mRNA sequence.
ACCESSION   BG618813
VERSION     BG618813.1  GI:13670184
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 750)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: k  column: 22
            High quality sequence stop: 622.
FEATURES             Location/Qualifiers
     source          1. .750
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767885"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      219 a    146 c    171 g    213 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.44e-95       Length:         750
Score:                  927.50         Matches:        190
Percent Similarity:     95.02%         Conservative:   1
Best Local Similarity:  94.53%         Mismatches:     6
Query Match:            48.53%         Indels:         5
DB:                     10             Gaps:           0

US-09-980-881A-2 (1-360) x BG618813 (1-750)

Qy          1 MetLysLeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisVal 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         11 ATGAAGCTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTC 70

Qy         21 PheAlaPheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGln 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 TTCGCGTTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAA 130

Qy         41 ValLeuGlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAsp 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 GTTCTACAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGAC 190

Qy         61 LeuIleValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 CTTATTGTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTG 250

Qy         81 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 AAAGCCCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGAT 310

Qy        101 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyr 120
              ||||||||||||||||||||||||||||||||    ||||||||||||||||||||||||
Db        311 CTTATTCAACAGCAGATTTCCAACGACACAGTCAGACCCCGAGCCTCCGCATCGTACTAT 370

Qy        121 GluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHis 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        371 GAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCAT 430

Qy        141 ProAspMet-LeuThrLysIleHisIleGly-SerSerPheGluLysTyrProLeu-Tyr 159
              |||||||| |||||||||||||||||||| |||||||||||||||   |||::: |||
Db        431 CCTGATATGCTTAACAAAAATCCACATTGGCATCCTCATTCGAGAAGTAACCAGTCGTAT 490

Qy        160 ValLeuLysValSer-GlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGl 179
              |||||||||||    ||||||||||       ||||||||||||||||||||||||||
Db        491 GTTTTAAAGGTTCCTGGGAAAAGAACAAGCAGGCAAAAATGCCATATGGATTGACTGTGG 550

Qy        179 yIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 197
              ||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        551 AATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTC-ATAGGCCAT 604

RESULT 14
CB157619
LOCUS       CB157619                 541 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0216807 L17N670205n1 Homo sapiens cDNA clone
            L17N670205n1-12-A11 5', mRNA sequence.
ACCESSION   CB157619
VERSION     CB157619.1  GI:28142751
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 12  row: A  column: 11
            High quality sequence stop: 541.
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L17N670205n1-12-A11"
                     /sex="F"
                     /lab_host="Top10F'"
                     /clone_lib="L17N670205n1"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
BASE COUNT      156 a    121 c    107 g    157 t
ORIGIN

Alignment Scores:
```

```
Pred. No.:              4.79e-95        Length:         541
Score:                  921.00          Matches:        178
Percent Similarity:     99.44%          Conservative:   0
Best Local Similarity:  99.44%          Mismatches:     1
Query Match:            48.19%          Indels:         0
DB:                     14              Gaps:           0

US-09-980-881A-2 (1-360) x CB157619 (1-541)

Qy          3 LeuCysSerLeuAlaValLeuValProIleValLeuPheCysGluGlnHisValPheAla 22
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          3 CTTTGCAGCCTTGCAGTCCTTGTACCCATTGTTCTCTTCTGTGAGCAGCATGTCTTCGCG 62

Qy         23 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 42
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         63 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 122

Qy         43 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 62
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        123 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 182

Qy         63 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 242

Qy         83 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 102
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 302

Qy        103 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 122
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 362

Qy        123 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 142
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 422

Qy        143 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 482

Qy        163 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 181
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db        483 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCAT 539

RESULT 15
BG565100
LOCUS       BG565100                747 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602583720F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711348 5',
            mRNA sequence.
ACCESSION   BG565100
VERSION     BG565100.1  GI:13572753
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1551  row: h  column: 05
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4711348"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      220 a    154 c    166 g    207 t
ORIGIN

Alignment Scores:
Pred. No.:              1.21e-92        Length:         747
Score:                  902.00          Matches:        194
Percent Similarity:     79.68%          Conservative:   6
Best Local Similarity:  77.29%          Mismatches:     9
Query Match:            47.20%          Indels:         45
DB:                     10              Gaps:           2

US-09-980-881A-2 (1-360) x BG565100 (1-747)

Qy         52 ValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPhePhe 71
              ||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db          1 GTTCTCTGGCAGCCGGTAACAGCTGACCTTAT-GTGAAGAAAAAACAAGTCCATTTTTTT 59

Qy         72 ValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCys 91
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 GTAGATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 119

Qy         92 SerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrVal 111
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 AGTGTCTTGCTGGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 179

Qy        112 SerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSer 131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 AGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCT 239

Qy        132 TrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySer 151
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 299

Qy        152 SerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 171
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        300 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAA 359

Qy        172 AsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCys 191
              ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db        360 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGACCTCTCCTGCTTTCTGC 419

Qy        192 LeuTrpPheIleGlyHis---------------------------------------- 197
              ||||||||||||||||
Db        420 TTGTGGTTCATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTC 479

Qy        198 ----------------------------------------------------AsnArgMet 200
                                                                 :::
Db        480 CTGAGGCTTGTGGATTTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCA 539

Qy        201 TrpArg--------------------LysAsnArgSerPheTyrAlaAsnAsnHisCys 213
              |||:::                  :::||||||||||||||||||||||||||||||
Db        540 TGGAAACAAGAATCGGAATGTGGAGAACAGAACCGTTCTTTCTATGCGAACAATCATTGC 599

Qy        214 IleGlyThrAspLeuAsnSerAsnPheValSerLysHisTrpCysGluGlu-GlyAlaSe 233
              ||||||||||||||||    |||      ||||||||||||||||||||| ||||||||
Db        600 ATCGGAACAGACCTGAATAGGAACTT-GCTTCCAAACACTGGTGTGAGGAAGGGTGCATC 658
```

```
Qy        233 rSerSerSer-CysSerGlu-ThrTyr-CysGlyLeuTyrProGluSerGluProGluVa 252
              ||||||||| |||||||| |||||| ||||||||||||||||||||||||||||||
Db        659 CAGTTCCTCATTGCTCGGAAAACCTACTTGTGGACTTTATCCTGAGTCAGAACCAGAGTG 718

Qy        252 lLysAlaValAlaSerPheLeuArg 260
               :::|||||||||||||||||:::
Db        719 -GAAGCAGTGGCTAGTTTCTTGAAG 742
```

Search completed: January 2, 2004, 02:37:19
Job time : 3203.89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:00:56 ; Search time 58.9207 Seconds
(without alignments)
910.539 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1798	100.0	338	21	AAB11458	Human brain carbox
2	1785	99.3	360	21	AAB11457	Human brain carbox
3	1722	95.8	386	24	ABG72095	Human protease pro
4	1717	95.5	386	24	ABG72093	Human protease pro
5	1699.5	94.5	423	17	AAR90293	Human plasma carbo
6	1699.5	94.5	423	18	AAW14733	Human plasma carbo
7	1699.5	94.5	423	20	AAW92270	Human plasma carbo
8	1694.5	94.2	423	14	AAR36273	Human plasma carbo
9	1338	74.4	246	21	AAB11459	Human brain carbox
10	816.5	45.4	211	21	AAB58129	Lung cancer associ
11	646.5	36.0	404	16	AAR75131	Porcine Tyr-His-Me
12	637.5	35.5	415	17	AAW06172	Human pancreatic c
13	637.5	35.5	415	19	AAW74476	Human pancreatic c
14	637.5	35.5	417	24	ABB99688	Amino acid sequenc
15	637.5	35.5	424	17	AAW06175	ProHCPB with PelB
16	636.5	35.4	324	22	AAG66562	Bothops jararaca c
17	622	34.6	417	23	AAU84325	Protein CPA3 diffe
18	622	34.6	667	24	ABP97414	Anti-human seminal
19	619.5	34.5	716	18	AAW41414	PreproHCPB-linker-
20	612	34.0	434	24	ABB99687	Amino acid sequenc
21	612	34.0	437	22	AAE00502	Human carboxypepti
22	612	34.0	437	23	AAE19175	Human protease, PR
23	610	33.9	307	17	AAW00602	Rat carboxypeptida
24	582	32.4	306	16	AAR75132	Porcine carboxypep
25	582	32.4	306	22	AAU04477	Porcine carboxypep
26	569	31.6	354	22	AAG66560	Human secreted met
27	569	31.6	354	22	AAG66566	Human secreted met
28	569	31.6	362	22	AAB74682	Human protease and
29	569	31.6	374	22	AAG66547	Human secreted met
30	569	31.6	374	22	AAG66565	Human secreted met
31	569	31.6	374	23	AAU82703	Amino acid sequenc
32	569	31.6	374	24	ABJ37888	NOVX protein seque
33	569	31.6	374	24	ABG74658	Human zinc-binding
34	569	31.6	444	23	AAE24240	Human 23566 (carbo
35	567	31.5	329	17	AAW06174	Mature HCPB with P
36	567	31.5	349	17	AAW06173	mature HCPB-(His)6
37	560	31.1	349	17	AAW06181	Modified HCPB (D25
38	560	31.1	349	18	AAW13749	Carboxypeptidase B
39	559	31.1	349	17	AAW06182	Modified HCPB (D25
40	559	31.1	349	18	AAW13750	Carboxypeptidase B
41	556	30.9	349	18	AAW13763	Carboxypeptidase B
42	554	30.8	349	18	AAW13752	Carboxypeptidase B
43	554	30.8	349	18	AAW13758	Carboxypeptidase B
44	554	30.8	399	22	AAE01664	Novel human protea
45	553	30.8	349	18	AAW13751	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAB11458
ID AAB11458 standard; Protein; 338 AA.
XX
AC AAB11458;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
XX

PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Disclosure; Page 71-73; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 338 AA;

Query Match 100.0%; Score 1798; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 1.1e-179;
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 338
              ||||||||||||||||||||||||||||||||||||||
Db        301 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 338
```

RESULT 2
AAB11457
ID AAB11457 standard; Protein; 360 AA.
XX
AC AAB11457;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
DR N-PSDB; AAC81962.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Claim 1; Page 68-71; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 360 AA;

Query Match 99.3%; Score 1785; DB 21; Length 360;
Best Local Similarity 99.1%; Pred. No. 2.9e-178;
Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 202

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              |||||||||||||||| || ||||||||||||||||||||||||||||||||||||||||
Db        203 KNRSFYANNHCIGTDLNSNFVSKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIDKTSKNTRYTHGHGS 322

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 338
              ||||||||||||||||||||||||||||||||||||||
Db        323 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360
```

RESULT 3
ABG72095
ID ABG72095 standard; Protein; 386 AA.
XX
AC ABG72095;
XX
DT 14-FEB-2003 (first entry)
XX
DE Human protease protein, variation #2.
XX
KW Human; enzyme; protease; proteolytic degradation; proteolysis;
KW proliferation; differentiation; signalling; therapeutic; gene therapy;
KW protein therapy; diagnostic; immune response; vaccine; inflammation;
KW cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW single nucleotide polymorphism; SNP.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers

```
FT   Misc-difference 169
FT                   /note= "This residue is the minor variant of a single
FT                          nucleotide polymorphism. The major variant is an
FT                          Ala residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Page -; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #2, the gene for which is located on chromosome 13.
CC   Note: This sequence is not shown in the specification but has been
CC   created by the indexer from information given in figure 3.
XX
SQ   Sequence   386 AA;

  Query Match             95.8%;  Score 1722;  DB 24;  Length 386;
  Best Local Similarity   99.7%;  Pred. No. 1.4e-171;
  Matches  323;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 202

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGS 322

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFT 324
              |||||||||||||||||||||||
Db        323 ETLYLAPGGGDDWIYDLGIKYSFT 346


RESULT 4
ABG72093
ID   ABG72093 standard; Protein; 386 AA.
XX
AC   ABG72093;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #1.
XX
KW   Human; enzyme; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 169
FT                   /note= "This residue is the major variant of a single
FT                          nucleotide polymorphism. The minor variant is an
FT                          Thr residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Fig 2; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
```

```
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #1, the gene for which is located on chromosome 13.
XX
SQ   Sequence   386 AA;

  Query Match             95.5%;  Score 1717;  DB 24;  Length 386;
  Best Local Similarity   99.4%;  Pred. No. 4.5e-171;
  Matches  322;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 202

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGS 322

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFT 324
              |||||||||||||||||||||||
Db        323 ETLYLAPGGGDDWIYDLGIKYSFT 346


RESULT 5
AAR90293
ID   AAR90293 standard; Protein; 423 AA.
XX
AC   AAR90293;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active-site     181
FT                   /note= "Residue suspecting of being involved in
FT                          catalytic activity of enzyme."
FT   Binding-site    183
FT                   /note= "Residue suspected of being part of the
FT                          binding site of enzyme."
FT   Binding-site    256..257
FT                   /note= "Residues suspected of being part of the
FT                          binding site of enzyme."
FT   Active-site     310
FT                   /note= "Residue suspecting of being involved in
FT                          catalytic activity of enzyme."
FT   Binding-site    312
FT                   /note= "Residue suspected of being part of the
FT                          binding site of enzyme."
FT   Active-site     369
FT                   /note= "Residue suspecting of being involved in
FT                          catalytic activity of enzyme."
FT   Binding-site    370..371
FT                   /note= "Residues suspected of being part of the
FT                          binding site of enzyme."
FT   Binding-site    394
FT                   /note= "Residue suspected of being part of the
FT                          binding site of enzyme."
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   N-PSDB; AAT11671.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

  Query Match             94.5%;  Score 1699.5;  DB 17;  Length 423;
  Best Local Similarity   89.8%;  Pred. No. 3.6e-169;
  Matches  324;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                           ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
                |
Db        383 T 383


RESULT 6
AAW14733
ID   AAW14733 standard; Protein; 423 AA.
XX
AC   AAW14733;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "Signal peptide"
FT   Protein         23..423
FT                   /note= "Mature PCPB"
FT   Cleavage-site   114
FT                   /note= "Potential clip site for activation of PCPB
FT                         as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    183
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     184
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    256
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    257
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     310
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    312
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     369
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    370
FT                   /note= "Expected to be involved in substrate
FT                         binding, thought to determine specificity
FT                         of PCPB as a carboxypeptidase B"
FT   Binding-site    371
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    394
FT                   /note= "Expected to be involved in substrate binding"
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   N-PSDB; AAT62846.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence represents human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

  Query Match             94.5%;  Score 1699.5;  DB 18;  Length 423;
  Best Local Similarity   89.8%;  Pred. No. 3.6e-169;
  Matches  324;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                           ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
                |
Db        383 T 383


RESULT 7
AAW92270
ID   AAW92270 standard; Protein; 423 AA.
XX
AC   AAW92270;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW   polymorphism detection; thrombotic disease.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "signal peptide"
```

```
FT   Protein         23..423
FT                   /note= "mature PCPBthr147"
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   N-PSDB; AAV74302.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Disclosure; Page 25-26; 35pp; English.
XX
CC   This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence   423 AA;

  Query Match             94.5%;  Score 1699.5;  DB 20;  Length 423;
  Best Local Similarity   89.8%;  Pred. No. 3.6e-169;
  Matches  324;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

```
RESULT 8
AAR36273
ID   AAR36273 standard; Protein; 423 AA.
XX
AC   AAR36273;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /label= signal_peptide
FT   Cleavage-site   114
FT                   /note= "putative site of cleavage by trypsin to
FT                          activate PCPB as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    183
FT                   /note= "expected to be involved in substrate
FT                          binding"
FT   Active-site     184
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    256..257
FT                   /note= "expected to be involved in substrate
FT                          binding"
FT   Active-site     310
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    312
FT                   /note= "expected to be involved in substrate
FT                          binding"
FT   Active-site     369
FT                   /note= "expected to be involved in catalytic
FT                          activity"
FT   Binding-site    370..371
FT                   /note= "expected to be involved in substrate
FT                          binding; Asp 370 is the residue which
FT                          determines specificity of the PCPB as
FT                          a carboxypeptidase B"
FT   Binding-site    394
FT                   /note= "expected to be involved in substrate
FT                          binding"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   N-PSDB; AAQ41001.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Claim 2; Fig 4; 40pp; English.
XX
CC   This amino acid sequence was deduced from a clone isolated from a
```

```
CC   human liver cDNA library. There is about 40% sequence identity
CC   between the deduced preproPCPB amino acid sequence and pro-human
CC   mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC   carboxypeptidase B. Human PCPB has the same substrate binding sites
CC   as, and shares 6 cysteine residues which form 3 intramolecular
CC   disulphide bonds with, bovine and rat carboxypeptidase B. The
CC   presence of the same amino acid (Asp at position 348 of the mature
CC   protein) at the region in carboxypeptidases that determines
CC   substrate specificity, suggests that PCPB represents a plasma-
CC   derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC   tPA of plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence   423 AA;

  Query Match            94.2%;  Score 1694.5;  DB 14;  Length 423;
  Best Local Similarity  89.5%;  Pred. No. 1.2e-168;
  Matches  323;  Conservative    0;  Mismatches    1;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 --------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                               ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGACSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383


RESULT 9
AAB11459
ID   AAB11459 standard; Protein; 246 AA.
XX
AC   AAB11459;
XX
DT   01-MAR-2001  (first entry)
XX
DE   Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
KW   Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW   treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW   cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW   hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW   Down's syndrome; head trauma.
XX
OS   Homo sapiens.
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Disclosure; Page 74-75; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   246 AA;

  Query Match            74.4%;  Score 1338;  DB 21;  Length 246;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-131;
  Matches  246;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         93 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60

Qy        153 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120

Qy        213 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180

Qy        273 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240

Qy        333 LPLSLK 338
              ||||||
Db        241 LPLSLK 246


RESULT 10
AAB58129
ID   AAB58129 standard; Protein; 211 AA.
XX
AC   AAB58129;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
```

```
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18005.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 944-945; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   211 AA;

  Query Match             45.4%;  Score 816.5;  DB 21;  Length 211;
  Best Local Similarity   90.6%;  Pred. No. 5.7e-77;
  Matches  154;  Conservative    1;  Mismatches    6;  Indels    9;  Gaps    2;

Qy        163 WISPAFCLWFI--------GHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 214
              |||   | | :          ||||||||||||||||||||||||||||||||||||||
Db          3 WIS-MLCRWLMWMVMNYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 61

Qy        215 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 121

Qy        275 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 171


RESULT 11
AAR75131
ID   AAR75131 standard; Protein; 404 AA.
XX
AC   AAR75131;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine Tyr-His-Met Procarboxypeptidase B.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB.
XX
OS   Sus scrofa.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   N-PSDB; AAQ90600.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Example 1; Page 16-17; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence   404 AA;

  Query Match             36.0%;  Score 646.5;  DB 16;  Length 404;
  Best Local Similarity   38.0%;  Pred. No. 1.1e-58;
  Matches  137;  Conservative   60;  Mismatches  121;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|           : :|  | :| :|  |:| :   |      | | | | |:  |:
Db         10 FEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 69

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||: ::  ::: |  :    | :   ||:|::   | :| : :|  :||
Db         70 FLEQNELQYEVLINNLRSVLEAQFDSRV---RTTGHSYEKYNNWETIEAWTKQVTSENPD 126

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::|    :|:||| ||    | ||::||| ||||||| ||| ||:
Db        127 LISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCGFHAREWISHAFCQWFVREAVLTY 185

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                   |:      |||||| ||   |   ||||| ||||
Db        186 GYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGTTCIGTDPNRNF-DA 244

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||  |||:  | |||||   ||| | ||:| |:| |:: ||||:::|||| |::||||
Db        245 GWCTTGASTDPCDETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 304

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|  ||: :  |   |:||:| |: |:| | || ||| || ||||||
Db        305 DYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGATTIYPAAGGSDDWAYDQGIKYSF 363

Qy        324 T 324
              |
Db        364 T 364


RESULT 12
AAW06172
ID   AAW06172 standard; Protein; 415 AA.
XX
AC   AAW06172;
XX
DT   12-FEB-1997  (first entry)
XX
```

```
DE   Human pancreatic carboxypeptidase B.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..13
FT                   /note= "partial enzyme secretion leader sequence;
FT                         (pre-sequence)"
FT   Peptide         14..108
FT                   /label= pro-sequence
FT   Protein         109..415
FT                   /label= mature_enzyme
FT   Misc-difference 243
FT                   /note= "not observed in other published HPCB
FT                        sequences"
FT   Active-site     361
FT                   /note= "Asp253 characteristic of B type specificity"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   N-PSDB; AAT42494.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 14; Page 131-132; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is that of the native prepro-HCPB.
XX
SQ   Sequence   415 AA;

  Query Match             35.5%;  Score 637.5;  DB 17;  Length 415;
  Best Local Similarity   36.0%;  Pred. No. 1.1e-57;
  Matches  130;  Conservative   68;  Mismatches  120;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |: :
Db         20 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 79

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
                |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         80 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 136

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        137 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 195

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :| ||| || :  : ||||| ||||
Db        196 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 254

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
                ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||||| :::|||
Db        255 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 314

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        315 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 373

Qy        324 T 324
              |
Db        374 T 374
```

```
RESULT 13
AAW74476
ID   AAW74476 standard; Protein; 415 AA.
XX
AC   AAW74476;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Human pancreatic carboxypeptidase protein sequence.
XX
KW   Human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW   prodrug therapy.
XX
OS   Homo sapiens.
XX
PN   WO9835988-A1.
XX
PD   20-AUG-1998.
XX
PF   10-FEB-1998;   98WO-GB00415.
XX
PR   29-OCT-1997;   97GB-0022727.
PR   14-FEB-1997;   97GB-0003104.
PR   18-OCT-1997;   97GB-0022003.
XX
PA   (ZENE ) ZENEPA   (ZENE ) ZENECA LTD.
XX
PI   Edge MD;
XX
DR   WPI; 1998-467168/40.
DR   N-PSDB; AAV41795.
XX
PT   New modified pro-domain of carboxy-peptidase B - enhances expression
PT   of co-expressed proteins for production of recombinant
PT   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT   enzyme prodrug therapy
XX
PS   Example 1; Page 54-55; 83pp; English.
XX
CC   The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC   modified pro-domain of CPB on a separate gene to enhance recombinant
CC   expression.  This process can be used to produce recombinant CPB in
CC   eukaryotic cells, or fusions of CPB with antibody chains.  CPB is used in
CC   insulin production and protein sequencing, while its fusions with
CC   antibodies are useful in antibody-directed enzyme prodrug therapy.  The
CC   modified pro-domain provide increased yields of recombinant CPB, possibly
CC   by protecting the C-terminus against enzymatic degradation or by
CC   increasing intracellular trafficking.
XX
SQ   Sequence   415 AA;

  Query Match             35.5%;  Score 637.5;  DB 19;  Length 415;
  Best Local Similarity   36.0%;  Pred. No. 1.1e-57;
  Matches  130;  Conservative   68;  Mismatches  120;  Indels   43;  Gaps    6;
```

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |      | | | | |   |:
Db         20 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 79

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         80 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 136

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        137 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 195

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :| ||| || :  : ||||| ||||
Db        196 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 254

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::||||| :::|||| 
Db        255 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 314

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        315 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 373

Qy        324 T 324
              |
Db        374 T 374
```

RESULT 14
ABB99688
ID ABB99688 standard; Protein; 417 AA.
XX
AC ABB99688;
XX
DT 28-MAR-2003 (first entry)
XX
DE Amino acid sequence of human zinc carboxypeptidase B precursor.
XX
KW Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW chronic obstructive pulmonary disease; COPD; haematological disorder;
KW zinc carboxypeptidase B.
XX
OS Homo sapiens.
XX
PN WO200295018-A1.
XX
PD 28-NOV-2002.
XX
PF 16-MAY-2002; 2002WO-EP05392.
XX
PR 22-MAY-2001; 2001US-292329P.
XX
PA (FARB) BAYER AG.
XX
PI Smolyar A;
XX
DR WPI; 2003-120801/11.
XX
PT New polynucleotide encoding a zinc carboxypeptidase B-like protein
PT polypeptide, useful for treating diseases associated with
PT carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT haematological disorder -
XX
PS Disclosure; Fig 3; 113pp; English.
XX
CC The present sequence represents a zinc carboxypeptidase B precursor.
CC The specification describes a human zinc carboxypeptidase B-like
CC protein. The zinc carboxypeptidase B-like protein is used for treating
CC diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
CC disease (COPD) or haematological disorders. The zinc carboxypeptidase
CC B-like protein may also be used to identify compounds which may act
CC as activators or inhibitors at the enzyme's active site, to raise
CC specific antibodies which can block the enzyme and effectively reduce
CC its activity, as a bait protein in a two-hybrid or three-hybrid assay
CC to identify other proteins which bind to or interact with the human
CC zinc carboxypeptidase-like protein polypeptide and modulate its
CC activity, and for immunization of mammals.
XX
SQ Sequence 417 AA;

Query Match 35.5%; Score 637.5; DB 24; Length 417;
Best Local Similarity 36.0%; Pred. No. 1.1e-57;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |      | | | | |   |:
Db         22 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         82 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 138

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 197

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :| ||| || :  : ||||| ||||
Db        198 GREIQVTELLNKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 256

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::||||| :::|||| 
Db        257 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 316

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        317 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 375

Qy        324 T 324
              |
Db        376 T 376
```

RESULT 15
AAW06175
ID AAW06175 standard; Protein; 424 AA.
XX
AC AAW06175;
XX
DT 17-FEB-1997 (first entry)
XX
DE ProHCPB with PelB leader sequence.
XX
KW ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW Fd; F(ab')2; PelB; leader; human carboxypeptidase B.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..22
FT /label= PelB_leader_sequence
FT Peptide 23..117
FT /label= pro_sequence
FT Protein 118..424
FT /label= mature_HCPB
XX
PN WO9620011-A1.
XX
PD 04-JUL-1996.
XX
PF 21-DEC-1995; 95WO-GB02991.
XX

PR 16-AUG-1995; 95GB-0016810.
PR 23-DEC-1994; 94GB-0026192.
XX
PA (ZENE) ZENECA LTD.
XX
PI Blakey DC, Boyle FT, Davies DH, Eggelte HJ, Heaton DW;
PI Hennam JF, Hennequin LFA, Marsham PR, Rabin BR, Slater AM;
PI Tarragona-Fiol A, Taylorson CJ;
XX
DR WPI; 1996-321650/32.
DR N-PSDB; AAT42506.
XX
PT Two component system for anti-tumour therapy - comprising targeting
PT moiety linked to mutated enzyme which can transform an
PT anti:neoplastic prodrug
XX
PS Reference Example 18; Page 140-142; 182pp; English.
XX
CC A two-component system for anti-tumour therapy comprises a targeting
CC moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC prodrug. The system is based on antibody directed enzyme prodrug therapy
CC (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC be used. The present sequence is that of proHCPB with a PelB leader
CC sequence encoded by pICI1738.
XX
SQ Sequence 424 AA;

Query Match 35.5%; Score 637.5; DB 17; Length 424;
Best Local Similarity 36.0%; Pred. No. 1.1e-57;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |:
Db         29 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 88

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         89 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 145

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        146 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 204

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                   |:      :| ||| || :  : |||| ||||
Db        205 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 263

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | ||||| ||| | ||:| |:|  :: ||||:::|||| :::||||
Db        264 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 323

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        324 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

Search completed: January 1, 2004, 19:28:06
Job time : 59.9207 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:21:46 ; Search time 20.8163 Seconds
(without alignments)
687.014 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1717	95.5	386	4	US-09-813-133A-2	Sequence 2, Appli
2	1699.5	94.5	423	1	US-07-649-591B-3	Sequence 3, Appli
3	1699.5	94.5	423	1	US-08-277-540-3	Sequence 3, Appli
4	1699.5	94.5	423	1	US-08-430-787A-3	Sequence 3, Appli
5	1699.5	94.5	423	2	US-08-869-057-2	Sequence 2, Appli
6	1688.5	93.9	423	4	US-09-813-133A-4	Sequence 4, Appli
7	646.5	36.0	404	1	US-08-696-139-2	Sequence 2, Appli
8	637.5	35.5	415	2	US-08-860-882A-57	Sequence 57, Appl
9	637.5	35.5	415	4	US-09-011-769A-39	Sequence 39, Appl
10	637.5	35.5	424	4	US-09-011-769A-56	Sequence 56, Appl
11	632	35.2	417	1	US-07-649-591B-7	Sequence 7, Appli
12	632	35.2	417	1	US-08-277-540-7	Sequence 7, Appli
13	632	35.2	417	1	US-08-430-787A-7	Sequence 7, Appli
14	622	34.6	417	1	US-07-649-591B-6	Sequence 6, Appli
15	622	34.6	417	1	US-08-277-540-6	Sequence 6, Appli
16	622	34.6	417	1	US-08-430-787A-6	Sequence 6, Appli
17	619.5	34.5	716	3	US-09-171-945-125	Sequence 125, App
18	612	34.0	437	4	US-09-675-305-10	Sequence 10, Appl
19	610	33.9	307	2	US-08-782-760-6	Sequence 6, Appli
20	610	33.9	307	5	PCT-US96-00995-6	Sequence 6, Appli
21	609.5	33.9	396	1	US-07-649-591B-4	Sequence 4, Appli
22	609.5	33.9	396	1	US-08-277-540-4	Sequence 4, Appli
23	609.5	33.9	396	1	US-08-430-787A-4	Sequence 4, Appli
24	582	32.4	306	1	US-08-696-139-4	Sequence 4, Appli
25	567	31.5	329	4	US-09-011-769A-51	Sequence 51, Appl
26	567	31.5	349	4	US-09-011-769A-47	Sequence 47, Appl
27	560	31.1	349	4	US-09-011-769A-60	Sequence 60, Appl
28	559	31.1	349	4	US-09-011-769A-64	Sequence 64, Appl
29	554	30.8	399	4	US-09-710-099-8	Sequence 8, Appli
30	549.5	30.6	417	1	US-07-649-591B-8	Sequence 8, Appli
31	549.5	30.6	417	1	US-08-277-540-8	Sequence 8, Appli
32	549.5	30.6	417	1	US-08-430-787A-8	Sequence 8, Appli
33	549	30.5	613	3	US-09-171-945-113	Sequence 113, App
34	534.5	29.7	350	4	US-09-675-305-12	Sequence 12, Appl
35	532.5	29.6	419	1	US-07-649-591B-5	Sequence 5, Appli
36	532.5	29.6	419	1	US-08-277-540-5	Sequence 5, Appli
37	532.5	29.6	419	1	US-08-430-787A-5	Sequence 5, Appli
38	527.5	29.3	436	4	US-09-710-099-6	Sequence 6, Appli
39	516.5	28.7	417	3	US-08-640-906-4	Sequence 4, Appli
40	516.5	28.7	417	4	US-09-395-936-4	Sequence 4, Appli
41	515.5	28.7	417	3	US-08-640-906-18	Sequence 18, Appl
42	515.5	28.7	417	4	US-09-395-936-18	Sequence 18, Appl
43	491.5	27.3	419	3	US-08-640-906-2	Sequence 2, Appli
44	491.5	27.3	419	4	US-09-395-936-2	Sequence 2, Appli
45	486.5	27.1	419	3	US-08-640-906-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Query Match 95.5%; Score 1717; DB 4; Length 386;
Best Local Similarity 99.4%; Pred. No. 1.5e-176;
Matches 322; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 202

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGS 322

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFT 324
              ||||||||||||||||||||||||
Db        323 ETLYLAPGGGDDWIYDLGIKYSFT 346
```

RESULT 2

```
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-3

  Query Match             94.5%;  Score 1699.5;  DB 1;  Length 423;
  Best Local Similarity   89.8%;  Pred. No. 1.3e-174;
  Matches  324;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

```
RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-3

  Query Match             94.5%;  Score 1699.5;  DB 1;  Length 423;
  Best Local Similarity   89.8%;  Pred. No. 1.3e-174;
  Matches  324;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322
```

```
Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

RESULT 4
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
; GENERAL INFORMATION:
; APPLICANT: Drayna, Dennis T., Eaton, Dan L.
; TITLE OF INVENTION: No. 5593674el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430,787A
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/277,540
; FILING DATE: 19-JUL-1994
; APPLICATION NUMBER: 08/167727
; FILING DATE: 15-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/959944
; FILING DATE: 14-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649591
; FILING DATE: 01-FEB-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Hasak, Janet E.
; REGISTRATION NUMBER: 28,616
; REFERENCE/DOCKET NUMBER: 689D1C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1896
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-430-787A-3

Query Match 94.5%; Score 1699.5; DB 1; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 -----------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                          ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
; GENERAL INFORMATION:
; APPLICANT: Morser, Michael J
; APPLICANT: Nagashima, Mariko
; TITLE OF INVENTION: Method of Detecting Thrombotic Disease
; TITLE OF INVENTION: Risk
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Berlex Biosciences Legal Department
; STREET: 15049 San Pablo Avenue
; CITY: Richmond
; STATE: California
; COUNTRY: USA
; ZIP: 94804-0099
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/869,057
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Washtien, Wendy L
; REGISTRATION NUMBER: 36,301
; REFERENCE/DOCKET NUMBER: 51509AUSM1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-262-5411
; TELEFAX: 510-262-7095
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; TISSUE TYPE: Plasma
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 23..401
US-08-869-057-2

Query Match 94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.3e-174;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82
```

```
Qy          61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy         121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy         176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||||||||||||||||||
Db         203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy         204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy         264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy         324 T 324
               |
Db         383 T 383
```

```
RESULT 6
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001173
;  CURRENT APPLICATION NUMBER: US/09/813,133A
;  CURRENT FILING DATE:  2001-06-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Human
US-09-813-133A-4

  Query Match            93.9%;  Score 1688.5;  DB 4;  Length 423;
  Best Local Similarity  89.2%;  Pred. No. 2e-173;
  Matches  322;  Conservative    0;  Mismatches    2;  Indels   37;  Gaps    1;

Qy           1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy          61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy         121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
               |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy         176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||||||||||||||||||
Db         203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy         204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy         264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
               |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db         323 TRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy         324 T 324
               |
Db         383 T 383
```

```
RESULT 7
US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  404 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-696-139-2

  Query Match            36.0%;  Score 646.5;  DB 1;  Length 404;
  Best Local Similarity  38.0%;  Pred. No. 4.7e-61;
  Matches  137;  Conservative   60;  Mismatches  121;  Indels   43;  Gaps    6;

Qy           1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
               |:  :|          : :|  | :| :|  |:| :   |     | | | | |:  |:
Db          10 FEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 69

Qy          61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||: ::  ::: |  :     | :   ||:|::   | :| : :|  :||
Db          70 FLEQNELQYEVLINNLRSVLEAQFDSRV---RTTGHSYEKYNNWETIEAWTKQVTSENPD 126

Qy         121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
               ::::  ||::|    :|:||| ||    | ||::||| ||||||| ||| ||:
Db         127 LISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCGFHAREWISHAFCQWFVREAVLTY 185

Qy         174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                    |:      |||||| ||  |   ||||| ||||
Db         186 GYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGTTCIGTDPNRNF-DA 244
```

```
Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||  |||:  | |||||   ||| | ||:| |:| |:: ||||:::|||||  |::|||| 
Db        245 GWCTTGASTDPCDETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 304

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|  ||: :  |   |:||:| |: |:| | || ||| || ||||||
Db        305 DYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGATTIYPAAGGSDDWAYDQGIKYSF 363

Qy        324 T 324
              |
Db        364 T 364
```

RESULT 8
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
; APPLICANT: TAYLORSON, CHRISTOPHER JOHN
; APPLICANT: EGGELTE, HENDRIKUS JOHANNES
; APPLICANT: TARRAGONA-FIOL, ANTONIO
; APPLICANT: RABIN, BRIAN ROBERT
; APPLICANT: BOYLE, FRANCIS THOMAS
; APPLICANT: HENNAM,JOHN FREDERICK
; APPLICANT: BLAKELY, DAVID CHARLES
; APPLICANT: MARSHAM, PETER ROBERT
; APPLICANT: HEATON, DAVID WILLIAM
; APPLICANT: DAVIES, DAVID HUW
; TITLE OF INVENTION: CHEMICAL COMPOUNDS
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PILLSBURY, MADISON & SUTRO
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/860,882A
; FILING DATE: JUNE 23, 1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: DONALD J. BIRD
; REGISTRATION NUMBER: 25,323
; REFERENCE/DOCKET NUMBER: 9901/238653
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3027
; TELEFAX: (202) 822-0944
; TELEX: 6174627 CUSH
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-860-882A-57

Query Match 35.5%; Score 637.5; DB 2; Length 415;
Best Local Similarity 36.0%; Pred. No. 4.6e-60;
Matches 130; Conservative 68; Mismatches 120; Indels 43; Gaps 6;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |: 
Db         20 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 79

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
                 |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
```

```
Db         80 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 136

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||     | ||::||| |||||||||| ||:
Db        137 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 195

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                   |:      :| ||| || :  : ||||| ||||
Db        196 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 254

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | |||||  ||| | ||:| |:|  :: ||||:::||||| :::|||| 
Db        255 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 314

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        315 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 373

Qy        324 T 324
              |
Db        374 T 374
```

RESULT 9
US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
; GENERAL INFORMATION:
; APPLICANT: SLATER, Anthony M.
; BLAKEY, David C.
; DAVIES, David H.
; HENNAM, John F.
; HENNEQUIN, Laurent F.A.
; MARSHAM, Peter R.
; DOWELL, Robert I.
; TITLE OF INVENTION: Chemical Compounds
; NUMBER OF SEQUENCES: 87
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pillsbury Madison & Sutro, LLP
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 1.44 Mb disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/011,769A
; FILING DATE: 13-Feb-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/01975
; FILING DATE: 13-AUG-1996
; APPLICATION NUMBER: GB 9612295.7
; FILING DATE: 12-JUN-1996
; APPLICATION NUMBER: GB 9611019.2
; FILING DATE: 25-MAY-1996
; APPLICATION NUMBER: GB 9516810.0
; FILING DATE: 16-AUG-1995
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 415 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

Query Match 35.5%; Score 637.5; DB 4; Length 415;

```
  Best Local Similarity   36.0%;  Pred. No. 4.6e-60;
  Matches  130;  Conservative   68;  Mismatches  120;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |:
Db         20 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 79

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
                |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         80 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 136

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        137 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 195

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      :| ||| || :  : ||||| ||||
Db        196 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 254

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||| :::||||
Db        255 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 314

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        315 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 373

Qy        324 T 324
              |
Db        374 T 374


RESULT 10
US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;    INFORMATION FOR SEQ ID NO: 56:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 424 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56

  Query Match             35.5%;  Score 637.5;  DB 4;  Length 424;
  Best Local Similarity   36.0%;  Pred. No. 4.7e-60;
  Matches  130;  Conservative   68;  Mismatches  120;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |:
Db         29 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 88

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
                |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         89 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 145

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        146 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 204

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      :| ||| || :  : ||||| ||||
Db        205 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 263

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||| :::||||
Db        264 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 323

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        324 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 382

Qy        324 T 324
              |
Db        383 T 383


RESULT 11
US-07-649-591B-7
; Sequence 7, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
```

```
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-7

  Query Match             35.2%;  Score 632;  DB 1;  Length 417;
  Best Local Similarity   38.2%;  Pred. No. 1.8e-59;
  Matches  138;  Conservative   57;  Mismatches  122;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|     :  :   ||:||| : |:  | |      |      | | |:  :   :::
Db         21 FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVNMTVDFRVSEKESQTIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |     |   :|: |::: |::|    | :: | |   | :|:  ::| || | : |:||
Db         81 TLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYNDWDKIVSWTEKMLEKHP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: ||:   : ||::|||||||||||||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHAREWISPAFCQWFVYQATKS 196

Qy        174 -GHN-----------------------------RMWRKNRSFYANNHCIGTDLNRNFAS 202
               | |                             ||||||||   |: ||||||||||
Db        197 YGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSRNQNSTCIGTDLNRNF-D 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
               |      ::  |   | |  |||| | ||| :|:| ::| |||||: ||||| :: ||
Db        256 VSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLIPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||      :|::|  ||  |  |:  |   ||| :|  : |:|   |   ||:||||||::
Db        316 YTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 12
US-08-277-540-7
; Sequence 7, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-7

  Query Match             35.2%;  Score 632;  DB 1;  Length 417;
  Best Local Similarity   38.2%;  Pred. No. 1.8e-59;
  Matches  138;  Conservative   57;  Mismatches  122;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|     :  :   ||:||| : |:  | |      |      | | |:  :   :::
Db         21 FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVNMTVDFRVSEKESQTIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |     |   :|: |::: |::|    | :: | |   | :|:  ::| || | : |:||
Db         81 TLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYNDWDKIVSWTEKMLEKHP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: ||:   : ||::|||||||||||||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHAREWISPAFCQWFVYQATKS 196

Qy        174 -GHN-----------------------------RMWRKNRSFYANNHCIGTDLNRNFAS 202
               | |                             ||||||||   |: ||||||||||
Db        197 YGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSRNQNSTCIGTDLNRNF-D 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
               |      ::  |   | |  |||| | ||| :|:| ::| |||||: ||||| :: ||
Db        256 VSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLIPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||      :|::|  ||  |  |:  |   ||| :|  : |:|   |   ||:||||||::
Db        316 YTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 13
US-08-430-787A-7
; Sequence 7, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
```

```
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-7

  Query Match             35.2%;  Score 632;  DB 1;  Length 417;
  Best Local Similarity   38.2%;  Pred. No. 1.8e-59;
  Matches  138;  Conservative   57;  Mismatches  122;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |  :|    :  :   ||:||| : |:  | |     |     | | |:  :   :::
Db         21 FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVNMTVDFRVSEKESQTIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |    |   :|: |::: |::|    | :: | |   | :|:  ::| || | : |:||
Db         81 TLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYNDWDKIVSWTEKMLEKHP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: ||:   : ||::|||||||||||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHAREWISPAFCQWFVYQATKS 196

Qy        174 -GHN-----------------------------RMWRKNRSFYANNHCIGTDLNRNFAS 202
               | |                             ||||||||   |: ||||||||||
Db        197 YGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSRNQNSTCIGTDLNRNF-D 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |      ::  |   | |  |||| | ||| :|:| ::| ||||||: ||||| :: ||
Db        256 VSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLIPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||      :|::|  ||  |  |:  |   ||| :|  : |:|    |   ||:||||||::
Db        316 YTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 14
US-07-649-591B-6
; Sequence 6, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-6

  Query Match             34.6%;  Score 622;  DB 1;  Length 417;
  Best Local Similarity   35.5%;  Pred. No. 2.2e-58;
  Matches  128;  Conservative   71;  Mismatches  118;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|    |:  :|  ::::|  | |:  | |     :    | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::||||||||:||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 -------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                      |:      |||||||||   |: ||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |      ::  |:: | |  |||| | ||| :|:| ::|:|| ||: ||||| ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||      :||:|: ||      :  |   ||| :|    |:|   |   || |||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 15
US-08-277-540-6
; Sequence 6, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
```

```
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  417 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-6

 Query Match             34.6%;  Score 622;  DB 1;  Length 417;
 Best Local Similarity   35.5%;  Pred. No. 2.2e-58;
 Matches  128;  Conservative   71;  Mismatches  118;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|    |:  :|  ::::|  | |:  | |     :     | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
              |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::||||||||:||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                     |:      |||||||||   |: |||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |     ::  |:: | |  |||| | ||| :|:| ::|:|| ||: |||||  ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :||:|: ||       :  |   ||| :|    |:|   |   || ||||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


Search completed: January  1, 2004, 19:33:10
Job time : 21.8163 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:32:17 ; Search time 48.3361 Seconds
(without alignments)
1403.024 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1699.5	94.5	423	12	US-10-379-836-17	Sequence 17, Appl
2	1615.5	89.8	423	12	US-10-379-836-2	Sequence 2, Appli
3	1487.5	82.7	422	12	US-10-379-836-18	Sequence 18, Appl
4	1466.5	81.6	422	12	US-10-379-836-16	Sequence 16, Appl
5	816.5	45.4	211	9	US-09-925-302-467	Sequence 467, App
6	637.5	35.5	402	12	US-10-379-836-20	Sequence 20, Appl
7	622	34.6	417	12	US-10-341-434-188	Sequence 188, App
8	622	34.6	417	15	US-10-229-546-2	Sequence 2, Appli
9	622	34.6	417	15	US-10-229-546-9	Sequence 9, Appli
10	619.5	34.5	716	10	US-09-910-059-125	Sequence 125, App
11	612	34.0	437	12	US-10-274-639-12	Sequence 12, Appl
12	612	34.0	437	14	US-10-200-344-10	Sequence 10, Appl
13	569	31.6	374	9	US-09-888-615-61	Sequence 61, Appl
14	569	31.6	444	16	US-10-176-306-74	Sequence 74, Appl
15	554	30.8	399	15	US-10-200-910-8	Sequence 8, Appli
16	549	30.5	613	10	US-09-910-059-113	Sequence 113, App
17	534.5	29.7	350	14	US-10-200-344-12	Sequence 12, Appl
18	527.5	29.3	436	15	US-10-200-910-6	Sequence 6, Appli
19	519.5	28.9	436	14	US-10-076-535-2	Sequence 2, Appli
20	511.5	28.4	417	9	US-09-923-779-150	Sequence 150, App
21	507.5	28.2	428	9	US-09-925-297-528	Sequence 528, App
22	501.5	27.9	298	12	US-10-379-836-19	Sequence 19, Appl
23	493.5	27.4	406	9	US-09-925-297-517	Sequence 517, App
24	493.5	27.4	419	12	US-10-345-680-65	Sequence 65, Appl
25	491	27.3	421	11	US-09-946-374-234	Sequence 234, App
26	491	27.3	421	12	US-10-015-387A-234	Sequence 234, App
27	491	27.3	421	12	US-10-006-130A-234	Sequence 234, App
28	491	27.3	421	12	US-10-199-672-308	Sequence 308, App
29	491	27.3	421	12	US-10-006-172A-234	Sequence 234, App
30	491	27.3	421	12	US-10-187-749-308	Sequence 308, App
31	491	27.3	421	12	US-10-194-457-308	Sequence 308, App
32	491	27.3	421	12	US-10-184-642-308	Sequence 308, App
33	491	27.3	421	12	US-10-196-747-308	Sequence 308, App
34	491	27.3	421	12	US-10-015-392A-234	Sequence 234, App
35	491	27.3	421	12	US-10-017-253A-234	Sequence 234, App
36	491	27.3	421	12	US-10-173-689-308	Sequence 308, App
37	491	27.3	421	12	US-10-173-690-308	Sequence 308, App
38	491	27.3	421	12	US-10-173-691-308	Sequence 308, App
39	491	27.3	421	12	US-10-173-692-308	Sequence 308, App
40	491	27.3	421	12	US-10-173-694-308	Sequence 308, App
41	491	27.3	421	12	US-10-173-698-308	Sequence 308, App
42	491	27.3	421	12	US-10-173-699-308	Sequence 308, App
43	491	27.3	421	12	US-10-173-707-308	Sequence 308, App
44	491	27.3	421	12	US-10-174-569-308	Sequence 308, App
45	491	27.3	421	12	US-10-174-583-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Query Match 94.5%; Score 1699.5; DB 12; Length 423;
Best Local Similarity 89.8%; Pred. No. 7.7e-166;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHPFVNASDVDNVKA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHPFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
```

```
                                                  ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383


RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Papio hamadryas
US-10-379-836-2

  Query Match            89.8%;  Score 1615.5;  DB 12;  Length 423;
  Best Local Similarity  85.0%;  Pred. No. 3.3e-157;
  Matches  307;  Conservative    8;  Mismatches    9;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||||||||||||||||||||||||||||||||| |||||||||:||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIEKKKQVHFFVNSSDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||| |||::||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIELITEKYPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              ||||||||||:||:||||||||||||||:||||||||||||||||||||||
Db        143 MLTKIHIGSSYEKHPLYVLKVSGKEQTAKNAMWIDCGIHAREWISPAFCLWFIGHITEYY 202

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||| ||||||||||||
Db        203 GIIGEYTNLLRHVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNRCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              |||||||| ||||||||||||||| ||||:||||||| ||||||||||||||||||||||
Db        263 HWCEEGASSFSCSETYCGLYPESEPEAKAVANFLRRNINHIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||:||||| ||||| |||||||||| ||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIQKTSKNIRYTHGRGSETLYLAPGGADDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383


RESULT 3
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-379-836-18

  Query Match            82.7%;  Score 1487.5;  DB 12;  Length 422;
  Best Local Similarity  78.1%;  Pred. No. 4.8e-144;
  Matches  282;  Conservative   20;  Mismatches   22;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||:||||||||||:||||||||:||||||| | |||:||||||||||:|||
Db         22 FQSGQVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKA 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||| || :||: :|||||:||  |||||||||||||||||||||||| |||:|||
Db         82 HLNVSRIPFNVLMNNVEDLIEQQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPD 141

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              || ||:|||||||||||||||  |||||||||||||||||||||:
Db        142 MLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFH 201

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||| : || |:||||||||||
Db        202 GKENLYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASK 261

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||:|||||||||||||||||||||||||| |||||: |||||||||||| |:|||||
Db        262 HWCEKGASSSSCSETYCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSY 321

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||  :|||||||| |||:||||||| ||||||||||||
Db        322 NRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSF 381

Qy        324 T 324
              |
Db        382 T 382


RESULT 4
US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-379-836-16

  Query Match            81.6%;  Score 1466.5;  DB 12;  Length 422;
```

```
  Best Local Similarity   76.7%;  Pred. No. 7e-142;
  Matches  277;  Conservative   22;  Mismatches   25;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||| ||:|||||||||||:||||||||:||||||||: | |||:||||||||||::|||
Db         22 FQSGHVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVNSVKA 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              :|| | || :||: :|||||||| ||||||||||:|||||||||||||||||| |||:|||
Db         82 YLNASRIPFNVLMNNVEDLIQQQTSNDTVSPRASSSYYEQYHSLNEIYSWIEVITEQHPD 141

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              || ||:||||:||||||||||||    ||||||||||||||||||||||||||:
Db        142 MLQKIYIGSSYEKYPLYVLKVSGKEHRVKNAIWIDCGIHAREWISPAFCLWFIGYVTQFH 201

Qy        176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                           ||||||||| : || |:||||||||||
Db        202 GKENTYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRKNRSVHMNNRCVGTDLNRNFASK 261

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||:|||| ||||||||||||||||||| ||||||| |||||||||||| |:|||||
Db        262 HWCEKGASSFSCSETYCGLYPESEPEVKAVADFLRRNINHIKAYISMHSYSQQILFPYSY 321

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              |||||||||||||||||||||  :|||||||| |||:|||||| ||||||||||||||||
Db        322 NRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSF 381

Qy        324 T 324
              |
Db        382 T 382


RESULT 5
US-09-925-302-467
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
;   LENGTH: 211
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-302-467

  Query Match            45.4%;  Score 816.5;  DB 9;  Length 211;
  Best Local Similarity   90.6%;  Pred. No. 1.7e-75;
  Matches  154;  Conservative    1;  Mismatches    6;  Indels    9;  Gaps    2;

Qy        163 WISPAFCLWFI--------GHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 214
              |||   | | :          ||||||||||||||||||||||||||||||||||||||
Db          3 WIS-MLCRWLMWMVMNYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 61

Qy        215 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 121

Qy        275 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 324
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 171


RESULT 6
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-379-836-20

  Query Match            35.5%;  Score 637.5;  DB 12;  Length 402;
  Best Local Similarity   36.0%;  Pred. No. 1.1e-56;
  Matches  130;  Conservative   68;  Mismatches  120;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |      | | | | |   |:
Db          7 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 66

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         67 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 123

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||     | ||::||| |||||||||| ||:
Db        124 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 182

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :| ||| || :  : ||||| ||||
Db        183 GREIQVTELLNKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 241

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||||  :::||||
Db        242 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 301

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        302 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 360

Qy        324 T 324
              |
Db        361 T 361


RESULT 7
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-341-434-188
```

```
  Query Match             34.6%;  Score 622;  DB 12;  Length 417;
  Best Local Similarity   35.5%;  Pred. No. 4.6e-55;
  Matches  128;  Conservative   71;  Mismatches  118;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|    |:  :|  ::::|  | |:  | |     :     | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::|||||||||:||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 -------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                        |:      |||||||||   |: ||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |     ::  |:: | |  |||| | ||| :|:| ::|:|| ||: ||||| ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :||:|: ||       :  |   ||| :|    |:|   |   || ||||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 8
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-2

  Query Match             34.6%;  Score 622;  DB 15;  Length 417;
  Best Local Similarity   35.5%;  Pred. No. 4.6e-55;
  Matches  128;  Conservative   71;  Mismatches  118;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|    |:  :|  ::::|  | |:  | |     :     | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::|||||||||:||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 -------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                        |:      |||||||||   |: ||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |     ::  |:: | |  |||| | ||| :|:| ::|:|| ||: ||||| ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :||:|: ||       :  |   ||| :|    |:|   |   || ||||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 9
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-9

  Query Match             34.6%;  Score 622;  DB 15;  Length 417;
  Best Local Similarity   35.5%;  Pred. No. 4.6e-55;
  Matches  128;  Conservative   71;  Mismatches  118;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|    |:  :|  ::::|  | |:  | |     :     | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::|||||||||:||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 -------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                        |:      |||||||||   |: ||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |     ::  |:: | |  |||| | ||| :|:| ::|:|| ||: ||||| ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :||:|: ||       :  |   ||| :|    |:|   |   || ||||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375
```

```
RESULT 10
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-125

  Query Match             34.5%;  Score 619.5;  DB 10;  Length 716;
  Best Local Similarity   35.2%;  Pred. No. 1.8e-54;
  Matches  127;  Conservative   69;  Mismatches  122;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : :::  | :| :|  |:| :   |      | | | | |   |:
Db         22 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :  ||:::: :::: |  :     ||:  ||:|:    | :| : :   :|
Db         82 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 138

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 197

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :| ||| || :  : ||||| ||||
Db        198 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 256

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||||| :::|||| 
Db        257 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 316

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|   |: :  :   |:||:| |: |:| : |   || ||.||:|||
Db        317 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPSAGTSKDWAYDQGIRYSF 375

Qy        324 T 324
              |
Db        376 T 376


RESULT 11
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 12
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

  Query Match             34.0%;  Score 612;  DB 12;  Length 437;
  Best Local Similarity   35.4%;  Pred. No. 5.2e-54;
  Matches  129;  Conservative   62;  Mismatches  127;  Indels   46;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK--KKQVHFFVNASDVDNV 58
              :   :|:  :|:|  :   |: ::   :: |||| :   : :     ||   | |    :
Db         37 YAGDKVIRFIPKTEEEAYALKKISYQLKVDLWQPSSISYVSEGTVTDVHIPQNGSRA--L 94

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS--YYEQYHSLNEIYSWIEFITE 116
              | |  : |   ||: |::  :::  |  |   | | |   || |||| || :|:  : :
Db         95 LAFLQEANIQYKVLIEDLQKTLEKGSSLHTQRNRRSLSGYNYEVYHSLEEIQNWMHHLNK 154

Qy        117 RHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI--- 173
              |  ::    || |:|   |::||: |:    | |:||||||||||| |||| ||:
Db        155 THSGLIHMFSIGRSYEGRSLFILKL-GRRSRLKRAVWIDCGIHAREWIGPAFCQWFVKEA 213

Qy        174 --------------------------------GHNRMWRKNRSFYANNHCIGTDLNRN 199
                                             ::| ||| ||  :   | | | |||
Db        214 LLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSWTNDRFWRKTRSRNSRFRCRGVDANRN 273

Qy        200 FASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVF 259
              :  | ||:||||   | :|||| :||||||||||:|||::   |:||:| |:|:| :::
Db        274 WKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAVANFLRKHRKHIRAYLSFHAYAQMLLY 332

Qy        260 PYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGI 319
              ||||  :   :   :   | :|| |::      || :|  | |||:: |   || |  ||
Db        333 PYSYKYATIPNFRCVESAAYKAVNALQSV-YGVRYRYGPASTTLYVSSGSSMDWAYKNGI 391

Qy        320 KYSF 323
              |:|
```

```
Db        392 PYAF 395


RESULT 12
US-10-200-344-10
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-200-344-10

  Query Match             34.0%;  Score 612;  DB 14;  Length 437;
  Best Local Similarity   35.4%;  Pred. No. 5.2e-54;
  Matches  129;  Conservative   62;  Mismatches  127;  Indels   46;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK--KKQVHFFVNASDVDNV 58
              :   :|:  :|:|  :   |: ::   :: |||| :   : :     ||   | |    :
Db         37 YAGDKVIRFIPKTEEEAYALKKISYQLKVDLWQPSSISYVSEGTVTDVHIPQNGSRA--L 94

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS--YYEQYHSLNEIYSWIEFITE 116
               | |  : |   ||: |::  :::  |  |   | | |   || |||| || :|:  : :
Db         95 LAFLQEANIQYKVLIEDLQKTLEKGSSLHTQRNRRSLSGYNYEVYHSLEEIQNWMHHLNK 154

Qy        117 RHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI--- 173
               |  ::    || |:|   |::||: |:    | |:||||||||||| |||| ||:
Db        155 THSGLIHMFSIGRSYEGRSLFILKL-GRRSRLKRAVWIDCGIHAREWIGPAFCQWFVKEA 213

Qy        174 ---------------------------------GHNRMWRKNRSFYANNHCIGTDLNRN 199
                                                ::| ||| ||  :   | | | |||
Db        214 LLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSWTNDRFWRKTRSRNSRFRCRGVDANRN 273

Qy        200 FASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVF 259
              :  | ||:||||   | :|||| :||||||||||:|||::   |:||:| |:|:| :::
Db        274 WKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAVANFLRKHRKHIRAYLSFHAYAQMLLY 332

Qy        260 PYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGI 319
              ||||  :   :   :  | :|| |::      || :|  | |||:: |   || |  ||
Db        333 PYSYKYATIPNFRCVESAAYKAVNALQSV-YGVRYRYGPASTTLYVSSGSSMDWAYKNGI 391

Qy        320 KYSF 323
               |:|
Db        392 PYAF 395


RESULT 13
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-888-615-61

  Query Match             31.6%;  Score 569;  DB 9;  Length 374;
  Best Local Similarity   39.0%;  Pred. No. 1.1e-49;
  Matches  110;  Conservative   51;  Mismatches   81;  Indels   40;  Gaps    5;

Qy         81 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVL 139
              :|:| : :||| :  :| |  || : ||| |:  |:|:: :::|:  :| ::| :|:| |
Db         29 RQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL 88

Qy        140 KVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI---------------------------- 173
              |:|      |  ||:||||||||||:|||| ||:
Db         89 KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLP 148

Qy        174 -----GH------NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGL 222
                   |:      :|:|||:|| : |  | |||||||| :  ||  ||| :   :|:||
Db        149 VLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGT 207

Qy        223 YPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAV 282
              | |||| ||||||:    : |  :::|||| | |: || ||::|| :| |:  |  :|
Db        208 GPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAA 267

Qy        283 RAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 324
              |: |    | |  |  :: || : |   ||  |:|| :|:|
Db        268 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 308


RESULT 14
US-10-176-306-74
; Sequence 74, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
```

```
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 444
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-306-74

  Query Match             31.6%;  Score 569;  DB 16;  Length 444;
  Best Local Similarity   39.0%;  Pred. No. 1.4e-49;
  Matches  110;  Conservative   51;  Mismatches   81;  Indels   40;  Gaps    5;

Qy         81 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVL 139
              :|:| : :||| :  :| |  || : ||| |:  |:|:: :::|:  :| ::| :|:| |
Db        159 RQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYL 218

Qy        140 KVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI---------------------------- 173
              |:|      |  ||:||||||||||:|||| ||:
Db        219 KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLP 278

Qy        174 -----GH------NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGL 222
                  |:      :|:|||:|| : |  | |||||||| :  ||  ||| :   :|:||
Db        279 VLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGT 337

Qy        223 YPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAV 282
              | |||| ||||||:    : |  :::|||| | |: || ||::|| :| |:  |  :|
Db        338 GPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAA 397

Qy        283 RAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 324
              |: |    | |  |  :: || : |   ||  |:|| :|:|
Db        398 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 438


RESULT 15
US-10-200-910-8
; Sequence 8, Application US/10200910
; Publication No. US20030009021A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Hilbun, Erin
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20030009021A1el Human Proteases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0086-USA
;  CURRENT APPLICATION NUMBER: US/10/200,910
;  CURRENT FILING DATE:  2002-07-22
;  PRIOR APPLICATION NUMBER: US/09/710,099
;  PRIOR FILING DATE: 2000-11-10
;  PRIOR APPLICATION NUMBER: US 60/165,260
;  PRIOR FILING DATE: 1999-11-12
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 399
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-200-910-8

  Query Match             30.8%;  Score 554;  DB 15;  Length 399;
  Best Local Similarity   39.0%;  Pred. No. 4.2e-48;
  Matches  130;  Conservative   60;  Mismatches  123;  Indels   20;  Gaps   10;

Qy          1 FQSGQVLAALPRTSRQVQVLQNL--TTTYEIVLWQ-PVTADLIVKKKQVHFFVNASDVDN 57
              |   |||  | :  :|: :| :|      ::  |: |    | |  :     |  |:: :
Db         37 FTGDQVLRVLAKDEKQLSLLGDLEGLKPQKVDFWRGPARPSLPVDMR-----VPFSELKD 91

Qy         58 VKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEF 113
              :||:|   |:  |::: |:: |:   :| ::       |:: |: |  ||:| ||||||:
Db         92 IKAYLESHGLAYSIMIKDIQVLLDEERQAMAKSRRLERSTNSFSYSSYHTLEEIYSWIDN 151

Qy        114 ITERHPDMLTKIHIGSSFEKYPLYVLKVS--GKEQTAKNAIWIDCGIHAREWISPAFCLW 171
                | |:::|| ||:|||   : ||| |  |       ||||| |||:||||: |  :|
Db        152 FVMEHSDIVSKIQIGNSFENQSILVLKFSTGGSRHP---AIWIDTGIHSREWITHATGIW 208

Qy        172 FIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVK 231
                ||:||||:|       ||| |||||: |  :   |::|: ||||| |  |:|||||
Db        209 TANKNRLWRKNKSIRPGIFCIGVDLNRNWKS-GFGGNGSNSNPCSETYHGPSPQSEPEVA 267

Qy        232 AVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKN 291
              |: :|:  : |  || ||:||||| :::||         :  ||  :| :|| |: |
Db        268 AIVNFITAHGN-FKALISIHSYSQMLMYPYGRLLEPVSNQRELYDLAKDAVEALYKV-HG 325

Qy        292 TRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 324
              |  |  | |||:| |   || || ||||:|:
Db        326 IEYIFGSISTTLYVASGITVDWAYDSGIKYAFS 358


Search completed: January  1, 2004, 19:50:58
Job time : 50.3361 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:25:56 ; Search time 23.286 Seconds
(without alignments)
1395.903 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1699.5	94.5	423	2	A41204	carboxypeptidase B
2	691.5	38.5	415	2	A32129	carboxypeptidase B
3	632	35.2	417	1	A34487	carboxypeptidase A
4	622	34.6	417	2	A43929	carboxypeptidase A
5	618	34.4	416	1	A42332	carboxypeptidase B
6	586	32.6	306	1	CPBOB	carboxypeptidase B
7	552	30.7	309	2	A38395	mast cell carboxyp
8	549.5	30.6	417	2	A32128	carboxypeptidase A
9	536.5	29.8	419	1	CPBOA	carboxypeptidase A
10	532.5	29.6	419	1	CPRTA	carboxypeptidase A
11	511.5	28.4	417	2	A56171	carboxypeptidase A
12	493.5	27.4	419	1	S29127	carboxypeptidase A
13	443	24.6	528	2	T33527	hypothetical prote
14	403	22.4	323	2	T20507	hypothetical prote
15	401	22.3	667	2	T33526	hypothetical prote
16	400.5	22.3	527	2	T27572	hypothetical prote
17	375.5	20.9	303	1	CPCYB	carboxypeptidase B
18	372.5	20.7	455	2	T33572	hypothetical prote
19	364	20.2	666	2	T24170	hypothetical prote
20	300	16.7	430	2	S48976	ECM14 protein - ye
21	286.5	15.9	424	2	S17571	carboxypeptidase T
22	283	15.7	451	2	S20723	carboxypeptidase (
23	252	14.0	497	2	T40260	carboxypeptidase p
24	250.5	13.9	373	2	T26030	hypothetical prote
25	184	10.2	999	2	T36021	probable zinc-bind
26	142.5	7.9	1446	2	T30916	carboxypeptidase D
27	140.5	7.8	558	2	S75104	hypothetical prote
28	116.5	6.5	707	2	T40070	origin recognition
29	109.5	6.1	985	2	T29910	hypothetical prote
30	108.5	6.0	985	2	D82776	pyruvate dehydroge
31	106.5	5.9	1389	2	I50090	carboxypeptidase g
32	105	5.8	491	2	B96739	hypothetical prote
33	101	5.6	663	2	T37772	telomere length re
34	101	5.6	1216	2	H84629	hypothetical prote
35	99	5.5	376	1	E69957	gamma-D-glutamyl-L
36	99	5.5	444	2	S51199	tryptophan 5-monoo
37	97.5	5.4	355	2	C83850	gamma-D-glutamyl-L
38	97	5.4	444	1	S10489	tryptophan 5-monoo
39	94	5.2	477	2	S16383	carboxypeptidase E
40	94	5.2	1160	2	I40589	parasporal crystal
41	94	5.2	1587	2	AB2012	hypothetical prote
42	93.5	5.2	318	2	S52424	homeodomain protei
43	93	5.2	1451	2	A36468	SPT6 protein - yea
44	92.5	5.1	986	2	E90596	restriction-modifi
45	92.5	5.1	986	2	H90565	restriction modifi

ALIGNMENTS

RESULT 1
A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N;Alternate names: plasma carboxypeptidase B
C;Species: Homo sapiens (man)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C;Accession: A41204
R;Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A;Title: Isolation, molecular cloning, and partial characterization of a novel carboxypep
A;Reference number: A41204; MUID:92042093; PMID:1939207
A;Accession: A41204
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <EAT>
A;Cross-references: GB:M75106; NID:g189686; PID:g189687
C;Genetics:
A;Gene: GDB:CPB2
A;Cross-references: GDB:129546; OMIM:212070
A;Map position: 13q14.11-13q14.11
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-114/Domain: activation peptide #status predicted <ACP>
F;115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Query Match 94.5%; Score 1699.5; DB 2; Length 423;
Best Local Similarity 89.8%; Pred. No. 1.1e-133;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 -------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                            ||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

RESULT 2
A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C;Accession: A32129; S17543
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: A32129
A;Molecule type: DNA
A;Residues: 1-415 <CLA>
A;Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R;Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A;Title: Distribution of manganese in rat pancreas and identification of its primary bin
A;Reference number: S17543; MUID:91378950; PMID:1898371
A;Accession: S17543
A;Molecule type: protein
A;Residues: 109-116,'X',118-130 <KOD>
C;Genetics:
A;Gene: CPB
A;Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;354,376/Active site: Tyr, Glu #status predicted

  Query Match   38.5%; Score 691.5; DB 2; Length 415;
  Best Local Similarity   40.2%; Pred. No. 9.1e-50;
  Matches 145; Conservative   59; Mismatches 114; Indels   43; Gaps   6;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|          | ::| | | ||  |:| :|  :      | | | | || :|:
Db         20 FDGNRVYRVSVHGEDHVNLIQELANTKEIDFWKPDSATQVKPLTTVDFHVKAEDVADVEN 79

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::| : :: |  : |   |||   | :|:    | :||: :   :||
Db         80 FLEENEVHYEVLISNVRNALESQFDSHT---RASGHSYTKYNKWETIEAWIQQVATDNPD 136

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::|:  ||::||   :||||: || :  | ||:|||| |||||||||| ||:
Db        137 LVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAIFIDCGFHAREWISPAFCQWFVREAVRTY 195

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :|||| ||  | : |:|   |||| :
Db        196 NQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRMWRKTRSTMAGSSCLGVRPNRNF-NA 254

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| | |||||||  |||| | ||:| |:| |:: ||||:::|||| :::||||
Db        255 GWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNLSTIKAYLTIHSYSQMMLYPYSY 314

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                :::|||: :   | : :  |   |:||:| |: |:| | || ||| || ||||||
Db        315 DYKLPENYEELNALVKGAAKEL-ATLHGTKYTYGPGATTIYPAAGGSDDWSYDQGIKYSF 373

Qy        324 T 324
              |
Db        374 T 374
```

RESULT 3
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34487
R;Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A;Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A;Reference number: A34487; MUID:90062123; PMID:2584208
A;Accession: A34487
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match   35.2%; Score 632; DB 1; Length 417;
  Best Local Similarity   38.2%; Pred. No. 8.2e-45;
  Matches 138; Conservative   57; Mismatches 122; Indels   44; Gaps   7;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|      :  :   ||:||| : |:  | |      |      | | |:  :   :::
Db         21 FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVNMTVDFRVSEKESQTIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |    |   :|: |::: |::|    | :: | |   | :|:  ::| || | : |:||
Db         81 TLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYNDWDKIVSWTEKMLEKHP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: ||:   : ||::|||||||||||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHAREWISPAFCQWFVYQATKS 196

Qy        174 -GHN----------------------------RMWRKNRSFYANNHCIGTDLNRNFAS 202
               | |                            ||||||||   |: ||||||||||
Db        197 YGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSRNQNSTCIGTDLNRNF-D 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |      ::  |   | |  |||| | ||| :|:| ::| |||||: ||||| :: ||
Db        256 VSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLIPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :|::|  ||  |  |:  |   ||| :|  : |:|    |   ||:||||||::
Db        316 YTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375
```

RESULT 4
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N;Alternate names: mast cell carboxypeptidase A3
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: A43929; A39246; A45759
R;Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A;Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A.
A;Reference number: A43929; MUID:92105393; PMID:1729276
A;Accession: A43929
A;Molecule type: DNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:M73716
A;Experimental source: mast cell
A;Note: the authors translated the codon CGC for residue 231 as Thr
R;Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafin, W
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A;Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison
A;Reference number: A39246; MUID:90083291; PMID:2594780
A;Accession: A39246
A;Molecule type: mRNA
A;Residues: 1-417 <RE2>
A;Cross-references: GB:M27717; NID:g179933; PIDN:AAA35652.1; PID:g179934
R;Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.
```
```

J. Clin. Invest. 83, 1630-1636, 1989
A;Title: Human mast cell carboxypeptidase. Purification and characterization.
A;Reference number: A45759; MUID:89214692; PMID:2708524
A;Accession: A45759
A;Molecule type: protein
A;Residues: 110-137 <GOL>
C;Genetics:
A;Gene: GDB:CPA3
A;Cross-references: GDB:125231; OMIM:114851
A;Map position: 3q21.3-3q25
A;Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-109/Domain: activation peptide #status predicted <ACT>
F;110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F;176,179,304/Binding site: zinc (His, Glu, His) #status predicted
F;245-268/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

Query Match 34.6%; Score 622; DB 2; Length 417;
Best Local Similarity 35.5%; Pred. No. 5.6e-44;
Matches 128; Conservative 71; Mismatches 118; Indels 44; Gaps 7;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |  :|    |:  :|  ::::|  | |:  | |      :     | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::|||||||||:||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 -------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                        |:      |||||||||   |: ||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |      ::  |:: | |  |||| | ||| :|:| ::|:|| ||: ||||| ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||      :||:|: ||      :  |   ||| :|    |:|   |   || ||||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375
```

RESULT 5
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N;Alternate names: carboxypeptidase B isozyme 1
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: A42332; S08254; S02813; S02812
R;Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992
A;Title: Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Ident
A;Reference number: A42332; MUID:92129345; PMID:1370825
A;Accession: A42332
A;Molecule type: mRNA
A;Residues: 1-416 <YAM>
A;Cross-references: GB:M81057; NID:g809194; PIDN:AAA66973.1; PID:g189625
A;Experimental source: pancreas
A;Note: sequence extracted from NCBI backbone (NCBIN:78573, NCBIP:78574)
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08254
A;Molecule type: protein
A;Residues: 16-40 <MOU>
A;Experimental source: pancreas
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02813
A;Molecule type: protein
A;Residues: 16-36,'X',38-43 <PAS>
A;Experimental source: pancreas
A;Accession: S02812
A;Molecule type: protein
A;Residues: 16-36,'Q',38-43 <PA2>
A;Experimental source: pancreas
C;Genetics:
A;Gene: GDB:CPB1
A;Cross-references: GDB:132160; OMIM:114852
A;Map position: 3q21.3-3q25
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-110/Domain: activation peptide #status predicted <ACP>
F;111-415/Product: carboxypeptidase B, pancreatic #status predicted <MAT>
F;173-186,258-272/Disulfide bonds: #status predicted
F;176,179,303/Binding site: zinc (His, Glu, His) #status predicted
F;355,377/Active site: Tyr, Glu #status predicted

Query Match 34.4%; Score 618; DB 1; Length 416;
Best Local Similarity 35.7%; Pred. No. 1.2e-43;
Matches 129; Conservative 68; Mismatches 120; Indels 44; Gaps 7;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|           : ::: | :| :|  |:| :   |     | | | | |   |:
Db         22 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::: :::: |  :      ||:   ||:|:    | :| : :   :|
Db         82 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 138

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 197

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      :| ||| || :  :  |||| ||||
Db        198 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGS-SIGTDPNRNF-DA 255

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::||||| :::||||
Db        256 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 315

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                ::: ||: :|    |: :  :  |:||:| |: |:| | || ||| || ||:|||
Db        316 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 374

Qy        324 T 324
              |
Db        375 T 375
```

RESULT 6
CPBOB
carboxypeptidase B (EC 3.4.17.2) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
C;Accession: A93797; A92150; A00912
R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A;Title: Amino-acid sequence of bovine carboxypeptidase B.
A;Reference number: A93797; MUID:75217824; PMID:1057162

```
A;Accession: A93797
A;Molecule type: protein
A;Residues: 1-306 <TIT>
R;Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations c
A;Reference number: A92150; MUID:74260705; PMID:4833744
A;Accession: A92150
A;Molecule type: protein
A;Residues: 31-93;131-181;263-265;292-306 <SCH>
R;Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A;Reference number: A92839; MUID:76265065; PMID:957425
A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R;Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptid
A;Reference number: A92051; MUID:70007159; PMID:5344132
A;Contents: annotation; active site
R;Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptida
A;Reference number: A92119; MUID:73061487; PMID:4565668
A;Contents: annotation; active site
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-158,149-163/Disulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

  Query Match             32.6%;  Score 586;  DB 1;  Length 306;
  Best Local Similarity   43.9%;  Pred. No. 3.7e-41;
  Matches  116;  Conservative   44;  Mismatches   64;  Indels   40;  Gaps    5;

Qy         98 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 157
              ||:|::   | :| | :   :||:::: ||::|    :|:||| ||  : | |:::|||
Db          6 YEKYNNWETIEAWTEQVASENPDLISRSAIGTTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy        158 IHAREWISPAFCLWFI------------------------------GH------NRMWR 180
              |||||||||||| ||:                              |:      |||||
Db         65 FHAREWISPAFCQWFVREAVRTYGREIHMTEFLDKLDFYVLPVVNIDGYIYTWTTNRMWR 124

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              | ||  | : | ||||||||     ||  |||:: |||||||   ||| | |||| |:| :
Db        125 KTRSTRAGSSCTGTDLNRNF-DAGWCSIGASNNPCSETYCGSAAESEKESKAVADFIRNH 183

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              :: ||||:::|||| :::||||     |:: ||: :|  ||: :  :   | |::| |:
Db        184 LSSIKAYLTIHSYSQMMLYPYSYDYKLPKNNVELNTLAKGAVKKL-ASLHGTTYSYGPGA 242

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFT 324
              |:| | || ||| || |||||||
Db        243 TTIYPASGGSDDWAYDQGIKYSFT 266


RESULT 7
A38395
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C;Accession: A38395; A33118
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A;Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme acti
A;Reference number: A38395; MUID:91105153; PMID:1988052
A;Accession: A38395
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-309 <COL>
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A;Reference number: A33118
A;Accession: A33118
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-309 <CO2>
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase
F;248,270/Active site: Tyr, Glu #status predicted

  Query Match             30.7%;  Score 552;  DB 2;  Length 309;
  Best Local Similarity   42.6%;  Pred. No. 2.5e-38;
  Matches  112;  Conservative   39;  Mismatches   72;  Indels   40;  Gaps    5;

Qy         98 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 157
              | :|:  |:| || | : |:||:|:::| |||: |  |||||| |::   : ||::|||
Db          8 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 66

Qy        158 IHAREWISPAFCLWFI-------GHN-----------------------------RMWR 180
              ||||||:|||| ||:       | |                             ||||
Db         67 IHAREWVSPAFCQWFVYQAAKSYGKNNIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 126

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||   :: ||||||||||     |      : : |   | |  |||| | ||| :|:| :
Db        127 KNRSKNPSSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 185

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              :| |||||: ||||| ::||| ||      :|::|  ||  |   :  :   ||| :|  :
Db        186 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy        301 ETLYLAPGGGDDWIYDLGIKYSF 323
              |:|   |   || ||||||::|
Db        245 STIYKTSGSSLDWAYDLGIKHTF 267


RESULT 8
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C;Accession: A32128
R;Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rutt
J. Biol. Chem. 263, 17828-17836, 1988
A;Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evo
A;Reference number: A32128; MUID:89034323; PMID:3182871
A;Accession: A32128
A;Molecule type: mRNA
A;Residues: 1-417 <GAR>
A;Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g5
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             30.6%;  Score 549.5;  DB 2;  Length 417;
  Best Local Similarity   34.6%;  Pred. No. 6.1e-38;
  Matches  134;  Conservative   63;  Mismatches  127;  Indels   63;  Gaps   12;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTT--YEIVLWQPVTADLIVKKKQVHFFVNASDVDNV 58
              |  ||| :|    |:: |  |      |: |:  |    :  : ||  |  : :  |
Db         20 FVGDQVLEIIPSHEEQIRTLLQLEAEEHLELDFWKSPT----IPGETVHVRVPFASIQAV 75

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISN---DTVSPRASASYYEQYHSLNEIYSWIEFIT 115
              |  |   ||  |::: ||: |: |:       :    |     :| ||:| |||  :: :
Db         76 KVFLESQGIDYSIMIEDVQVLLDQEREEMLFNQQRERGGNFNFEAYHTLEEIYQEMDNLV 135

Qy        116 ERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIHAREWISPAFCLW--- 171
              :| :::|:::||||  |: ||| | | :   | |||:| |||||||:: |  ||
Db        136 AENPGLVSKVNLGSSFENRPMNVLKFSTGGD---KPAIWLDAGIHAREWVTQATALWTAN 192

Qy        172 ---------------------FI-------GH------NRMWRKNRSFYANNHCIGTDLN 197
                                  |:       |:      |||||| ||  : : |:| | |
Db        193 KIASDYGTDPAITSLLNTLDIFLLPVTNPDGYVFSQTTNRMWRKTRSKRSGSGCVGVDPN 252

Qy        198 RNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHI 257
              ||:   ::   ||||| ||::| |  | || |||::  |:: :  ::||:|::||||| :
```

```
Db        253 RNW-DANFGGPGASSSPCSDSYHGPKPNSEVEVKSIVDFIKSH-GKVKAFITLHSYSQLL 310

Qy        258 VFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDL 317
              :||| |  :|  |  ||  || :| :|: |    | |  |     :| | ||  || |||
Db        311 MFPYGYKCTKPDDFNELDEVAQKAAQAL-KRLHGTSYKVGPICSVIYQASGGSIDWAYDL 369

Qy        318 GIKYSFTSN----------PPVEKLLP 334
              ||||||             | :::||
Db        370 GIKYSFAFELRDTAFYGFLLPAKQILP 396


RESULT 9
CPBOA
carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine
N;Alternate names: procarboxypeptidase A activation peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1992 #text_change 05-May-2000
C;Accession: JN0126; JT0440; A31406; S00059; A38834; A00910; A90355
R;Le Hueerou, I.; Guilloteau, P.; Toullec, R.; Puigserver, A.; Wicker, C.
Biochem. Biophys. Res. Commun. 175, 110-116, 1991
A;Title: Cloning and nucleotide sequence of a bovine pancreatic preprocarboxypeptidase A
A;Reference number: JN0126; MUID:91151335; PMID:1998496
A;Accession: JN0126
A;Molecule type: mRNA
A;Residues: 1-419 <LEH>
A;Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R;Foglizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A;Title: Primary structure of the activation peptide from bovine pancreatic procarboxype
A;Reference number: JT0440; MUID:88209583; PMID:3365436
A;Accession: JT0440
A;Molecule type: protein
A;Residues: 17-110 <FOG>
R;Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A;Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidas
A;Reference number: A31406; MUID:89150306; PMID:3147705
A;Accession: A31406
A;Molecule type: protein
A;Residues: 17-110 <WAD>
R;Chapus, C.; Kerfelec, B.; Foglizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A;Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A;Reference number: S00059; MUID:87275921; PMID:3609014
A;Accession: S00059
A;Molecule type: protein
A;Residues: 17-69 <CHA>
R;Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A;Title: The amino acid sequence of bovine carboxypeptidase A.
A;Reference number: A38834; MUID:70050818; PMID:5260942
A;Accession: A38834
A;Molecule type: protein
A;Residues: 111-137,'Q',139-140,'E',142-417 <BRA>
R;Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A;Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterizatic
A;Reference number: A90360; MUID:71155185; PMID:4927805
A;Accession: A00910
A;Molecule type: protein
A;Residues: 214-411 <BR2>
A;Note: this is the final paper in a series
R;Petra, P.H.; Hermodson, M.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 4023-4025, 1971
A;Title: Characterization of bovine carboxypeptidase A(Allan).
A;Reference number: A90355; MUID:72138789; PMID:5143102
A;Accession: A90355
A;Molecule type: protein
A;Residues: 111-142 <PET>
R;Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A;Title: Identification of the amino acid replacements characterizing the allotropic for
A;Reference number: A90556; MUID:69283620; PMID:5817619
A;Contents: annotation; allelic variant
A;Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CPA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137,'Q',139-14
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/Disulfide bonds: #status experimental
F;358,380/Active site: Tyr, Glu #status experimental

  Query Match             29.8%;  Score 536.5;  DB 1;  Length 419;
  Best Local Similarity   34.8%;  Pred. No. 7.4e-37;
  Matches  129;  Conservative   59;  Mismatches  124;  Indels   59;  Gaps   10;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--W----QPVTADLIVKKKQVHFFVNASD 54
              |   |||         :|| :: |     : |  |    ||           :   |
Db         20 FVGHQVLRITAADEAEVQTVKELEDLEHLQLDFWRGPGQP--------GSPIDVRVPFPS 71

Qy         55 VDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSW 110
               :  ||  |   ||   ::: ||: |:   |:|:       |:: :: |  ||:|:||| :
Db         72 LQAVKVFLEAHGIRYRIMIEDVQSLLDEEQEQMFASQSRARSTNTFNYATYHTLDEIYDF 131

Qy        111 IEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCL 170
              :: :   || :::|: || |:|  |:|||| |    : : ||||| |||:||||: | :
Db        132 MDLLVAEHPQLVSKLQIGRSYEGRPIYVLKFS-TGGSNRPAIWIDLGIHSREWITQATGV 190

Qy        171 WFI----------------------------------GHNRMWRKNRSFYANNHCIG 193
              ||                                  ||:||| ||  ::: |:|
Db        191 WFAKKFTEDYGQDPSFTAILDSMDIFLEIVTNPDGFAFTHSQNRLWRKTRSVTSSSLCVG 250

Qy        194 TDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSY 253
               | |||:    : : ||||| ||||| | |  || |||::  |:: : |  ||::|:|||
Db        251 VDANRNW-DAGFGKAGASSSPCSETYHGKYANSEVEVKSIVDFVKDHGN-FKAFLSIHSY 308

Qy        254 SQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDW 313
              || :::|| ||      |  ||: ||  || |: |:   | | :|    |:| | ||  ||
Db        309 SQLLLYPYGYTTQSIPDKTELNQVAKSAVEAL-KSLYGTSYKYGSIITTIYQASGGSIDW 367

Qy        314 IYDLGIKYSFT 324
               |: |||||||
Db        368 SYNQGIKYSFT 378


RESULT 10
CPRTA
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.L
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <QUI>
A;Cross-references: GB:J00713; NID:g203364; PIDN:AAA40893.1; PID:g203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Comparat
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA
```

```
A;Residues: 1-260,'F',262,'M',264-346,'K',348-419 <CLA>
C;Genetics:
A;Gene: CPA1
A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

  Query Match             29.6%;  Score 532.5;  DB 1;  Length 419;
  Best Local Similarity   35.3%;  Pred. No. 1.6e-36;
  Matches  131;  Conservative   57;  Mismatches  124;  Indels   59;  Gaps   10;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTADLIVKKKQVHFFVNASDVDNV 58
              |   |||         ||| :: |     : |  |:       |    :| |      : :|
Db         20 FVGHQVLRISAADEAQVQKVKELEDLEHLQLDFWRDAARAGIPIDVRVPF----PSIQSV 75

Qy         59 KAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFI 114
              || |    ||   ::: ||: |:   :||:|       :: |: |  ||:|:||| ::: :
Db         76 KAFLEYHGISYEIMIEDVQLLLDEEKQQMSAFQARALSTDSFNYATYHTLDEIYEFMDLL 135

Qy        115 TERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIG 174
                 || :::|| ||::||  |::||| |    | : ||||| |||:|||:: |  :||
Db        136 VAEHPQLVSKIQIGNTFEGRPIHVLKFS-TGGTNRPAIWIDTGIHSREWVTQASGVWFAK 194

Qy        175 H-----------------------------------NRMWRKNRSFYANNHCIGTDLN 197
                                                  |||||| ||    : |:| | |
Db        195 KVTKDYGQDPTFTAVLDNMDIFLEIVTNPDGFAYTHKTNRMWRKTRSHTQGSLCVGVDPN 254

Qy        198 RNFASKHW----CEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSY 253
              ||     |    : ||||: ||||| | :| || |||::  |:  : | |||:||:|||
Db        255 RN-----WDAGLGKAGASSNPCSETYRGKFPNSEVEVKSIVDFVTSHGN-IKAFISIHSY 308

Qy        254 SQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDW 313
              || :::|| ||   : |  ||  :|  || |:  :   | : :|   :|:| | |   ||
Db        309 SQLLLYPYGYTSEPAPDQAELDQLAKSAVTAL-TSLHGTEFKYGSIIDTIYQASGSTIDW 367

Qy        314 IYDLGIKYSFT 324
              |  |||||||
Db        368 TYSQGIKYSFT 378


RESULT 11
A56171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C;Accession: A56171; S02809; S71395
R;Catasus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAA74425.1; PID:g790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S',18-36,'N',38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996

A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LAE>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             28.4%;  Score 511.5;  DB 2;  Length 417;
  Best Local Similarity   33.1%;  Pred. No. 8.9e-35;
  Matches  128;  Conservative   61;  Mismatches  135;  Indels   63;  Gaps   10;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTADLIVKKKQVHFFVNASDVDNV 58
              |   |||  :|    |:: |  |      : |  |:  |       :  |  |   :|  |
Db         20 FVGDQVLEIVPSNEEQIKNLLQLEAQEHLQLDFWKSPT----TPGETAHVRVPFVNVQAV 75

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS---YYEQYHSLNEIYSWIEFIT 115
              |  |    ||  |::: ||: |: ::      :  |    |     :  ||:| ||   :: :
Db         76 KVFLESQGIAYSIMIEDVQVLLDKENEEMLFNRRRERSGNFNFGAYHTLEEISQEMDNLV 135

Qy        116 ERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIHAREWISPAFCLWFIG 174
                 || :::|::|||||  |: ||| | |  :   | |||:| |||||||:: |  ||
Db        136 AEHPGLVSKVNIGSSFENRPMNVLKFSTGGD---KPAIWLDAGIHAREWVTQATALWTAN 192

Qy        175 ------------------------------------HNRMWRKNRSFYANNHCIGTDLN 197
                                                 ||||||  ||  : : |:| | |
Db        193 KIVSDYGKDPSITSILDALDIFLLPVTNPDGYVFSQTKNRMWRKTRSKVSGSLCVGVDPN 252

Qy        198 RNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHI 257
              ||:    :   ||||: ||::| |     || |||::  |:: :  ::||:| :||||| :
Db        253 RNW-DAGFGGPGASSNPCSDSYHGPSANSEVEVKSIVDFIKSH-GKVKAFIILHSYSQLL 310

Qy        258 VFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDL 317
              :||| |  :|  | :||| || :| ::: ::   |:|  |     :| | ||  || ||
Db        311 MFPYGYKCTKLDDFDELSEVAQKAAQSL-RSLHGTKYKVGPICSVIYQASGGSIDWSYDY 369

Qy        318 GIKYSFTSN----------PPVEKLLP 334
              ||||||             |  ::||
Db        370 GIKYSFAFELRDTGRYGFLLPARQILP 396


RESULT 12
S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N;Alternate names: pancreatic carboxypeptidase A1
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: S29127; A34205; S08253; S02810; S71394; S02811
R;Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver,
Biochem. J. 287, 299-303, 1992
A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.
A;Reference number: S29127; MUID:93038569; PMID:1417781
A;Accession: S29127
A;Molecule type: mRNA
A;Residues: 1-419 <CAT>
A;Cross-references: EMBL:X67318; NID:g35329; PIDN:CAA47732.1; PID:g35330
R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
A;Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
```

```
A;Molecule type: DNA
A;Residues: 330-396 <STE>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43;'XXX',114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

  Query Match            27.4%;  Score 493.5;  DB 1;  Length 419;
  Best Local Similarity  33.1%;  Pred. No. 2.8e-33;
  Matches  126;  Conservative   58;  Mismatches  118;  Indels   79;  Gaps   12;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQ-------PVTADLIVKKKQVHFFVN 51
              |   |||         |||  :: |     : |  |:         |:          :| |
Db         20 FVGHQVLRISVADEAQVQKVKELEDLEHLQLDFWRGPAHPGSPIDV-------RVPF--- 69

Qy         52 ASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQI--------SNDTVSPRASASYYEQ 100
                 :  ||  |   ||    :: ||: |:   |:|:        | || :       |
Db         70 -PSIQAVKIFLESHGISYETMIEDVQSLLDEEQEQMFAFRSRARSTDTFN-------YAT 121

Qy        101 YHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHA 160
              ||:| ||| ::: :   :| :::|| ||:::|  |:|||| |    : : ||||| |||:
Db        122 YHTLEEIYDFLDLLVAENPHLVSKIQIGNTYEGRPIYVLKFS-TGGSKRPAIWIDTGIHS 180

Qy        161 REWISPAFCLWFI----------------------------------GHNRMWRKNR 183
              |||:: |  :||                                  |||||| |
Db        181 REWVTQASGVWFAKKITQDYGQDAAFTAILDTLDIFLEIVTNPDGFAFTHSTNRMWRKTR 240

Qy        184 SFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQ 243
              |  | : ||| | |||: :      ||||: ||||| | :  || |||::  |:: : |
Db        241 SHTAGSLCIGVDPNRNWDAGFGL-SGASSNPCSETYHGKFANSEVEVKSIVDFVKDHGN- 298

Qy        244 IKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETL 303
              |||:||:||||| :::|| |      | :||  ::  || |:  :   |:: :|   : :
Db        299 IKAFISIHSYSQLLMYPYGYKTEPVPDQDELDQLSKAAVTAL-ASLYGTKFNYGSIIKAI 357

Qy        304 YLAPGGGDDWIYDLGIKYSFT 324
              | | |   || |  |||||||
Db        358 YQASGSTIDWTYSQGIKYSFT 378


RESULT 13
T33527
hypothetical protein T06A4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C;Accession: T33527
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33527
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-528 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.1
A;Map position: 1
A;Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C;Superfamily: carboxypeptidase

  Query Match            24.6%;  Score 443;  DB 2;  Length 528;
  Best Local Similarity  32.6%;  Pred. No. 6.1e-29;
  Matches  121;  Conservative   57;  Mismatches  121;  Indels   72;  Gaps   14;

Qy         11 PRTSRQVQVLQNL---TTTYEIVLWQPVT-----ADLIVKKKQVHFFVNASDVDNVKAHL 62
              | |   |: |::|    : ||:  ||| |       || |        ||   |::: |
Db         54 PETEGSVKYLRSLYEDPSPYELDFWQPPTNIGAIVDLTVAPADAPRFV--KDLESKK--- 108

Qy         63 NVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDML 122
                  |   | : |:   |: :  :|       :   |::|:|| || : :: : : :| |:
Db        109 ----ISYIVAVNDLSKAIENERGSDKFYNPVAGFAYDKYNSLEEIQTEMKRLKKEYPTMI 164

Qy        123 TKIHIGSSFEKYPLYVLKVSGKEQT--AKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              | | || | |    | |:|::||     :| ::||| ||||||||:||  : :| |
Db        165 TLIDIGQSHENRTLLVMKITGKRNPLGSKISMWIDAGIHAREWIAPATAM-YIAHELLLG 223

Qy        176 --------------------------------NRMWRKNRS-------FYANNHCIGTD 195
                                             |||||||||       ::   | | |
Db        224 YENDATVAKLMDHIDFYILPVMNPDGYEYSREKNRMWRKNRSPAKCARQTFSTVCCSGVD 283

Qy        196 LNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASPLRRNINQIKAYISMHSYSQ 255
              |||||    |   |:||  | :|| |    |||| :||  || :|  :  |:||:|||||
Db        284 LNRNF-DWFWASTGSSSDPCHDTYHGSAAFSEPESQAVRDFLEQNTPE--AFISLHSYSQ 340

Qy        256 HIVFPYSYTR-SKSKD-HEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDW 313
              : || : : |  :| |  |  :|  | :|: :    |:|  | |:: :| | ||  ||
Db        341 MWLIPYGHRKQSYPQDYHTGLRPLALRATKALYEL-YGTKYQVGTGADLMYEASGGSHDW 399

Qy        314 IY-DLGIKYSF 323
              | : |::
Db        400 AKGQLKVPYAY 410


RESULT 14
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20507; T24289
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284
A;Accession: T20507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WIL>
A;Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone F02D8
R;Wild, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19869
A;Accession: T24289
A;Status: preliminary; translated from GB/EMBL/DDBJ
```

```
A;Molecule type: DNA
A;Residues: 1-323 <WI2>
A;Cross-references: EMBL:Z78413; PIDN:CAB01666.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone T01C3
C;Genetics:
A;Gene: CESP:F02D8.4
A;Map position: 5
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C;Superfamily: carboxypeptidase

  Query Match             22.4%;  Score 403;  DB 2;  Length 323;
  Best Local Similarity   33.5%;  Pred. No. 6.8e-26;
  Matches   89;  Conservative   52;  Mismatches   81;  Indels   44;  Gaps    7;

Qy        110 WIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFC 169
              ::  : :::|: :   :||:|:|   :  :::: : ::|  :||| |||||||||
Db          3 FLNSLAQQYPNDVKLQNIGNSYEGRSITAVRIA-DDGSSKPIVWIDAGIHAREWISYNVA 61

Qy        170 LWFI--------------------------------GHNRMWRKNRSFYANNHCIGTDL 196
              |: |                                ::||||| || : |: | | |
Db         62 LYLIYTIVSQPAYRNLLDSVQLVVVPNTNPDGYEYSRTNDRMWRKTRSRFTNSRCAGADA 121

Qy        197 NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQH 256
              |||:   :|  :| | | ||| :||  |:|||| |: : : |:  :|| ||::||| |
Db        122 NRNYPF-YWGTQGVSHSQCSEIFCGSRPQSEPEVLALTNAIIRDEERIKGYIALHSYGQE 180

Qy        257 IVFPYSYT-RSKSKDHEELSLVA---SEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDD 312
              |::|: :| |:   | ::|  |    : |:||:     || ||  :  : || | |  ||
Db        181 ILYPWGHTQRTYPTDVQDLIQVGRAMASAIRAV----NNTDYTVVNSGDGLYPAAGASDD 236

Qy        313 WIYDLGIKYSFTSN-PPVEKLLPLSL 337
              |    |||||:|    |::     ||
Db        237 WAKSRGIKYSYTIELSPIDDFTGFSL 262


RESULT 15
T33526
hypothetical protein T06A4.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T33526
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33526
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-667 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67474.1; GSPDB:GN00019; CESP:T06A4.3
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.3
A;Map position: 1
A;Introns: 50/3; 96/2; 167/3; 236/3; 276/3; 315/1; 365/3; 385/3; 444/1; 487/1; 582/3

  Query Match             22.3%;  Score 401;  DB 2;  Length 667;
  Best Local Similarity   27.9%;  Pred. No. 2.6e-25;
  Matches  119;  Conservative   62;  Mismatches  129;  Indels  116;  Gaps   14;

Qy          2 QSGQVLAALPRTSRQVQVLQNL---TTTYEIVLWQPVTA-----DLIVKKKQVHFFVNAS 53
              ||  |   ||::    | :| |    | :::  |:        |::|  |   :|::
Db         21 QSFAVYRLLPKSQTDFQAIQRLYKNATDHDLNFWKTGKDKHGFWDVMVDMKNSKYFLD-- 78

Qy         54 DVDNVKAHLNVSGIPCSVLLADVEDLIQQQISNDTVS---PR-----ASASY-YEQYHSL 104
                | |: |     : ||| || :  :| |:|   ||     :|| | :  | |
Db         79 -------FLQVNDISYIKTIDDVEGLITKHENNKTLSNMFPRLWDDSSSAHYDFHTYGSY 131

Qy        105 NEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWI 164
              :  |:: :  ::| |:  | || : |   :  ::: |  :| |   ||| |||||||
Db        132 QRMTDWMKQLVVKYPKMVQYISIGKTTEGRNIDGVEIGGDSRT-KKIFWIDGGIHAREWA 190

Qy        165 SPAFCLWFI---------------------------------------------------- 173
              :|   |:||
Db        191 APHTALFFIHQVCDPAPSEKSVFRLVPLGLHKCWQRQKKGKLDNKWLTSRANEPGIKKLL 250

Qy        174 ------------------------GHNRMWRKNRS--------FYANNHCIGTDLNRNFA 201
                                       | |:||||||        :  |  | | ||||||
Db        251 NEITFVVVPCLNPDGYEFTRSSTNPHVRLWRKNRSKMQCRKDIWGRNRCCRGVDLNRNF- 309

Qy        202 SKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFL--RRNINQIKAYISMHSYSQHIVF 259
              |: | | |   ||| | |  | |||| |||   |  :|   :  |||::|:|||  :
Db        310 DFHFRESGTSDDPCSEIYQGPSPFSEPEAKAVRDALLSQRYKGRTDAYITLHTYSQIWIH 369

Qy        260 PYSYTR-SKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWI-YDL 317
              || : : :   | ::|  |  :| :|:::    |:|  | |::||| | || :||  ::
Db        370 PYGHKKDAYPGDIKDLYEVGKKAAQALKRV-YGTKYVVGSGADTLYPASGGSEDWAKHEA 428

Qy        318 GIKYSF 323
              :|: :
Db        429 KVKFVY 434


Search completed: January  1, 2004, 19:34:31
Job time : 25.536 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:03:36 ; Search time 13.4071 Seconds
(without alignments)
1185.568 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	691.5	38.5	415	1	CBPB_RAT	P19223 rattus norv
2	637.5	35.5	417	1	CBPB_HUMAN	P15086 homo sapien
3	635.5	35.3	401	1	CBPB_PIG	P09955 sus scrofa
4	632	35.2	417	1	CBPC_MOUSE	P15089 mus musculu
5	630.5	35.1	416	1	CBPB_CANFA	P55261 canis famil
6	622	34.6	417	1	CBPC_HUMAN	P15088 homo sapien
7	586	32.6	306	1	CBPB_BOVIN	P00732 bos taurus
8	552	30.7	309	1	CBPC_RAT	P21961 rattus norv
9	549.5	30.6	417	1	CBP2_RAT	P19222 rattus norv
10	536.5	29.8	419	1	CBPA_BOVIN	P00730 bos taurus
11	533.5	29.7	419	1	CBP1_RAT	P00731 rattus norv
12	514.5	28.6	417	1	CPB2_HUMAN	P48052 homo sapien
13	493.5	27.4	419	1	CBP1_HUMAN	P15085 homo sapien
14	491	27.3	421	1	CBP4_HUMAN	Q9ui42 homo sapien
15	411.5	22.9	304	1	CBPZ_SIMVI	P42788 simulium vi
16	393	21.9	433	1	CBPA_ANOGA	O02350 anopheles g
17	375.5	20.9	303	1	CBPB_ASTFL	P04069 astacus flu
18	300	16.7	430	1	YHT2_YEAST	P38836 saccharomyc
19	286.5	15.9	424	1	CBPT_THEVU	P29068 thermoactin
20	283	15.7	451	1	CBPS_STRGR	P18143 streptomyce
21	259	14.4	434	1	CBPS_STRCP	P39041 streptomyce
22	122.5	6.8	1380	1	CBPD_HUMAN	O75976 homo sapien
23	118.5	6.6	1377	1	CBPD_MOUSE	O89001 mus musculu
24	116.5	6.5	707	1	ORC1_SCHPO	P54789 schizosacch
25	110.5	6.1	722	1	CPXM_MOUSE	Q9z100 mus musculu
26	109	6.1	734	1	CPXM_HUMAN	Q96sm3 homo sapien
27	103.5	5.8	445	1	TPH1_CHICK	P70080 gallus gall
28	101	5.6	663	1	TAZ1_SCHPO	P79005 schizosacch
29	99	5.5	376	1	YQGT_BACSU	P54497 bacillus su
30	99	5.5	444	1	TPH1_HUMAN	P17752 homo sapien
31	99	5.5	1255	1	PER2_HUMAN	O15055 homo sapien
32	97	5.4	297	1	Y103_HUMAN	Q15006 homo sapien
33	95	5.3	444	1	TPH1_RABIT	P17290 oryctolagus
34	94	5.2	1160	1	C8CA_BACTP	Q45706 bacillus th
35	93	5.2	1451	1	SPT6_YEAST	P23615 saccharomyc
36	92	5.1	454	1	CBPH_LOPAM	P37892 lophius ame
37	92	5.1	488	1	TY3H_ANGAN	O42091 anguilla an
38	91	5.1	262	1	YCJI_ECOLI	P51983 escherichia
39	90.5	5.0	319	1	UN30_CAEEL	P52906 caenorhabdi
40	90.5	5.0	434	1	CBPH_BOVIN	P04836 bos taurus
41	90.5	5.0	476	1	CBPH_RAT	P15087 rattus norv
42	90.5	5.0	3588	1	SRF1_BACSU	P27206 bacillus su
43	89.5	5.0	476	1	CBPH_MOUSE	Q00493 mus musculu
44	89	4.9	481	1	TPH_XENLA	Q92142 xenopus lae
45	89	4.9	575	1	ACEB_YEAST	Q12031 saccharomyc

ALIGNMENTS

RESULT 1
CBPB_RAT
ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M23959; AAA40872.1; -.
DR EMBL; M23947; AAA40872.1; JOINED.
DR EMBL; M23950; AAA40872.1; JOINED.
DR EMBL; M23952; AAA40872.1; JOINED.
DR EMBL; M23953; AAA40872.1; JOINED.
DR EMBL; M23954; AAA40872.1; JOINED.
DR PIR; A32129; A32129.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 13 BY SIMILARITY.
FT PROPEP 14 108 ACTIVATION PEPTIDE.
FT CHAIN 109 415 CARBOXYPEPTIDASE B.
FT METAL 174 174 ZINC (BY SIMILARITY).
FT METAL 177 177 ZINC (BY SIMILARITY).
FT METAL 302 302 ZINC (BY SIMILARITY).

```
FT   ACT_SITE    376    376       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    171    184       BY SIMILARITY.
FT   DISULFID    243    266       BY SIMILARITY.
FT   DISULFID    257    271       BY SIMILARITY.
SQ   SEQUENCE   415 AA;  47515 MW;  8EA06CCADE30B6F2 CRC64;

  Query Match           38.5%;  Score 691.5;  DB 1;  Length 415;
  Best Local Similarity   40.2%;  Pred. No. 5.3e-52;
  Matches  145;  Conservative   59;  Mismatches  114;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|          | ::| |  | ||  |:| :|  :      | | | | || :|:
Db         20 FDGNRVYRVSVHGEDHVNLIQELANTKEIDFWKPDSATQVKPLTTVDFHVKAEDVADVEN 79

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::| : :: |  : |   |||    | :|:    | :||: :   :||
Db         80 FLEENEVHYEVLISNVRNALESQFDSHT---RASGHSYTKYNKWETIEAWIQQVATDNPD 136

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::|:  ||::||   :|||: || :  | ||:|||| |||||||||| ||:
Db        137 LVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAIFIDCGFHAREWISPAFCQWFVREAVRTY 195

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      :|||| ||  | : |:|   |||| :
Db        196 NQEIHMKQLLDELDFYVLPVVNIDGYVYTWTKDRMWRKTRSTMAGSSCLGVRPNRNF-NA 254

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||| ||| | |||||||  |||| | ||:| |:| |:: ||||:::|||| :::|||
Db        255 GWCEVGASRSPCSETYCGPAPESEKETKALADFIRNNLSTIKAYLTIHSYSQMMLYPYSY 314

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                 :::|||: :   | : :  |   |:||:| |: |:| | || ||| || ||||||
Db        315 DYKLPENYEELNALVKGAAKEL-ATLHGTKYTYGPGATTIYPAAGGSDDWSYDQGIKYSF 373

Qy        324 T 324
              |
Db        374 T 374


RESULT 2
CBPB_HUMAN
ID   CBPB_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15086; O60834; Q96BQ8;
DT   01-APR-1990 (Rel. 14, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE   (PASP).
GN   CPB1 OR CPB OR PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC   TISSUE=Pancreas;
RX   MEDLINE=92129345; PubMed=1370825;
RA   Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA   French C.K.;
RT   "Isolation of a cDNA encoding a human serum marker for acute
RT   pancreatitis. Identification of pancreas-specific protein as
RT   pancreatic procarboxypeptidase B.";
RL   J. Biol. Chem. 267:2575-2581(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=98182241; PubMed=9524066;
RA   Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA   Aviles F.X.;
RT   "Comparative analysis of the sequences and three-dimensional models
RT   of human procarboxypeptidases A1, A2 and B.";
RL   Biol. Chem. 379:149-155(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordân H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 16-43.
RC   TISSUE=Pancreas;
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- TISSUE SPECIFICITY: Pancreas.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81057; AAA66973.1; -.
DR   EMBL; AJ224866; CAA12163.1; -.
DR   EMBL; BC015338; AAH15338.1; -.
DR   PDB; 1KWM; 28-AUG-02.
DR   MEROPS; M14.003; -.
DR   Genew; HGNC:2299; CPB1.
DR   MIM; 114852; -.
DR   GO; GO:0004180; F:carboxypeptidase activity; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL        1     15
FT   PROPEP       16    110       ACTIVATION PEPTIDE.
FT   CHAIN       111    417       CARBOXYPEPTIDASE B.
FT   METAL       176    176       ZINC (BY SIMILARITY).
FT   METAL       179    179       ZINC (BY SIMILARITY).
FT   METAL       304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE    378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    173    186       BY SIMILARITY.
FT   DISULFID    245    268       BY SIMILARITY.
FT   DISULFID    259    273       BY SIMILARITY.
FT   CONFLICT     16     16       H -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT     17     17       H -> Q (IN REF. 1; AA SEQUENCE).
```

```
FT   CONFLICT     37     37       H -> Q (IN REF. 4).
FT   CONFLICT    208    208       D -> N (IN REF. 2).
FT   CONFLICT    245    245       MISSING (IN REF. 1).
SQ   SEQUENCE   417 AA;  47367 MW;  EBBB98B27F5D5AF9 CRC64;

  Query Match           35.5%;  Score 637.5;  DB 1;  Length 417;
  Best Local Similarity   36.0%;  Pred. No. 2.4e-47;
  Matches  130;  Conservative   68;  Mismatches  120;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : ::: | :| :|  |:| :   |     | | | | |   |:
Db         22 FEGEKVFRVNVEDENHINIIRELASTTQIDFWKPDSVTQIKPHSTVDFRVKAEDTVTVEN 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||:::: :::: |  :     ||:   ||:|:    | :| : :   :|
Db         82 VLKQNELQYKVLISNLRNVVEAQFDSRV---RATGHSYEKYNKWETIEAWTQQVATENPA 138

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||    | ||::||| |||||||||| ||:
Db        139 LISRSVIGTTFEGRAIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFCQWFVREAVRTY 197

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                   |:      :| ||| || :  : |||| ||||
Db        198 GREIQVTELLDKLDFYVLPVLNIDGYIYTWTKSRFWRKTRSTHTGSSCIGTDPNRNF-DA 256

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||| ||| : | |||||   ||| | ||:| |:|  :: ||||:::|||| :::|||
Db        257 GWCEIGASRNPCDETYCGPAAESEKETKALADFIRNKLSSIKAYLTIHSYSQMMIYPYSY 316

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                 ::: ||: :|   |: :  :   |:||:| |: |:| | || ||| || ||:|||
Db        317 AYKLGENNAELNALAKATVKEL-ASLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIRYSF 375

Qy        324 T 324
              |
Db        376 T 376


RESULT 3
CBPB_PIG
ID   CBPB_PIG       STANDARD;      PRT;   401 AA.
AC   P09955;
DT   01-MAR-1989 (Rel. 10, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2).
GN   CPB.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE OF 1-103.
RX   MEDLINE=91208150; PubMed=2018774;
RA   Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA   Aviles F.X.;
RT   "Analysis of the activation process of porcine procarboxypeptidase B
RT   and determination of the sequence of its activation segment.";
RL   Biochemistry 30:4082-4089(1991).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 1-38.
RX   MEDLINE=85279427; PubMed=4026847;
RA   Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT   "Sequential homologies between procarboxypeptidases A and B from
RT   porcine pancreas.";
RL   Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=91114690; PubMed=1989878;
RA   Coll M., Guasch A., Aviles F.X., Huber R.;
RT   "Three-dimensional structure of porcine procarboxypeptidase B: a
RT   structural basis of its inactivity.";
RL   EMBO J. 10:1-9(1991).
RN   [4]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX   MEDLINE=91027767; PubMed=2223783;
RA   Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT   "Sequence-specific 1H NMR assignments and determination of the
RT   secondary structure for the activation domain isolated from
RT   pancreatic procarboxypeptidase B.";
RL   Biochemistry 29:7515-7522(1990).
RN   [5]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=91114693; PubMed=1989879;
RA   Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
RT   "The NMR structure of the activation domain isolated from porcine
RT   procarboxypeptidase B.";
RL   EMBO J. 10:11-15(1991).
RN   [6]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=93044373; PubMed=1422143;
RA   Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA   Huber R., Wuethrich K.;
RT   "Comparison of the NMR solution structure with the X-ray crystal
RT   structure of the activation domain from procarboxypeptidase B.";
RL   J. Biomol. NMR 2:1-10(1992).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR   PDB; 1PBA; 31-OCT-93.
DR   PDB; 1NSA; 24-DEC-97.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW   3D-structure.
FT   PROPEP        1     95       ACTIVATION PEPTIDE.
FT   CHAIN        96    401       CARBOXYPEPTIDASE B.
FT   DISULFID    158    171
FT   DISULFID    230    253
FT   DISULFID    244    258
FT   METAL       161    161       ZINC.
FT   METAL       164    164       ZINC.
FT   METAL       289    289       ZINC.
FT   ACT_SITE    341    341
FT   ACT_SITE    363    363       NUCLEOPHILE.
FT   TURN          8      9
FT   STRAND       11     17
FT   HELIX        20     32
FT   STRAND       36     39
FT   HELIX        43     45
FT   STRAND       50     56
FT   HELIX        58     60
FT   HELIX        61     69
FT   TURN         70     72
FT   STRAND       74     79
FT   HELIX        82     87
FT   TURN         88     88
FT   TURN        101    102
FT   HELIX       107    120
FT   TURN        122    124
FT   STRAND      125    132
FT   TURN        134    135
FT   STRAND      138    144
FT   TURN        149    150
FT   STRAND      153    157
FT   TURN        162    163
FT   HELIX       165    181
```

```
FT   TURN        182    184
FT   HELIX       186    194
FT   STRAND      196    200
FT   HELIX       205    213
FT   TURN        214    214
FT   TURN        216    217
FT   STRAND      224    224
FT   TURN        226    227
FT   STRAND      231    231
FT   HELIX       235    237
FT   TURN        243    244
FT   STRAND      249    249
FT   TURN        252    253
FT   TURN        255    256
FT   STRAND      257    257
FT   TURN        262    263
FT   HELIX       266    277
FT   TURN        278    281
FT   STRAND      282    289
FT   STRAND      294    297
FT   TURN        307    308
FT   HELIX       309    327
FT   STRAND      332    335
FT   HELIX       336    339
FT   HELIX       347    353
FT   TURN        354    355
FT   STRAND      359    363
FT   TURN        371    372
FT   HELIX       376    378
FT   HELIX       379    398
FT   TURN        399    401
SQ   SEQUENCE   401 AA;  45713 MW;  53129AF159A26348 CRC64;

  Query Match             35.3%;  Score 635.5;  DB 1;  Length 401;
  Best Local Similarity   37.4%;  Pred. No. 3.3e-47;
  Matches  135;  Conservative   64;  Mismatches  119;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|        :  |  | :| :|  |:| :   |      | | | | |:  |:
Db          7 FEGEKVFRVNVEDENDISELHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 66

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||: ::  ::: |  :     | :   ||:|::   | :| | :| ::||
Db         67 FLEQNELQYEVLINNLRSVLEAQFDSRC---RTTGHSYEKYNNWETIEAWTEQVTSKNPD 123

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::|:   :|:||| ||  :  | ||::||| ||||||| ||| ||:
Db        124 LISRSAIGTTFDGDNIYLLKV-GKPGSNKPAIFMDCGFHAREWISQAFCQWFVRDAVRTY 182

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      |||||| ||   | : | ||| |||| :
Db        183 GYEAHMTEFLDNLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGSSCTGTDPNRNF-NA 241

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||  ||| : |:|||||   ||| | ||:| |:| |:: ||||:::||||| |::||||
Db        242 GWCTVGASVNPCNETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 301

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::  ||: :|  ||: :  :    | |::| || |:| | || ||| |: ||||||
Db        302 DYKLPENDAELNSLAKGAVKEL-ASLYGTSYSYGPGSTTIYPAAGGSDDWAYNQGIKYSF 360

Qy        324 T 324
              |
Db        361 T 361


RESULT 4
CBPC_MOUSE
ID   CBPC_MOUSE     STANDARD;      PRT;   417 AA.
AC   P15089;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90062123; PubMed=2584208;
RA   Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA   Serafin W.E.;
RT   "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT   novel member of the carboxypeptidase gene family.";
RL   J. Biol. Chem. 264:20094-20099(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05118; AAA37369.1; -.
DR   PIR; A34487; A34487.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   MGD; MGI:88479; Cpa3.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     15
FT   PROPEP       16    109       ACTIVATION PEPTIDE.
FT   CHAIN       110    417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL       176    176       ZINC (BY SIMILARITY).
FT   METAL       179    179       ZINC (BY SIMILARITY).
FT   METAL       304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE    378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    173    186       BY SIMILARITY.
FT   DISULFID    245    268       BY SIMILARITY.
SQ   SEQUENCE   417 AA;  48790 MW;  A2B300A068D1BA6D CRC64;

  Query Match             35.2%;  Score 632;  DB 1;  Length 417;
  Best Local Similarity   38.2%;  Pred. No. 7e-47;
  Matches  138;  Conservative   57;  Mismatches  122;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|      :  :   ||:||| : |:  | |      |      | | |:  :   :::
Db         21 FDREKVFRVKLQNEKHASVLKNLTQSIELDFWYPDAIHDIAVNMTVDFRVSEKESQTIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
                |    |   :|: |::: |::|    | :: | |   | :|:  ::| || | : |:||
Db         81 TLEQHKIHYEILIHDLQEEIEKQFDVKDEIAGRHS---YAKYNDWDKIVSWTEKMLEKHP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  |||||: ||:   : ||::|||||||||||||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCGIHAREWISPAFCQWFVYQATKS 196

Qy        174 -GHN----------------------------RMWRKNRSFYANNHCIGTDLNRNFAS 202
               | |                           ||||||||   |: ||||||||||
Db        197 YGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWRKNRSRNQNSTCIGTDLNRNF-D 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
```

```
                |     ::  |   | |  |||| | ||| :|:| ::| |||||: ||||| :: ||
Db        256 VSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLIPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :|::|  ||  |  |:  |   ||| :|  : |:|   |   ||:||||||::
Db        316 YTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIASTIYKTSGSSLDWVYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375


RESULT 5
CBPB_CANFA
ID   CBPB_CANFA     STANDARD;      PRT;   416 AA.
AC   P55261;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE   membrane associated protein) (ZAP47).
GN   CPB1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Fukuoka S.-I.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D78348; BAA11366.1; -.
DR   HSSP; P09955; 1PBA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     15       BY SIMILARITY.
FT   PROPEP      16    109       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN      110    416       CARBOXYPEPTIDASE B.
FT   METAL      175    175       ZINC (BY SIMILARITY).
FT   METAL      178    178       ZINC (BY SIMILARITY).
FT   METAL      303    303       ZINC (BY SIMILARITY).
FT   ACT_SITE   377    377       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   244    267       BY SIMILARITY.
FT   DISULFID   258    272       BY SIMILARITY.
SQ   SEQUENCE   416 AA;  47595 MW;  DFE1992CD52F8AB4 CRC64;

  Query Match             35.1%;  Score 630.5;  DB 1;  Length 416;
  Best Local Similarity   36.6%;  Pred. No. 9.4e-47;
  Matches  132;  Conservative   64;  Mismatches  122;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|          : :|  | :| :|  |:| :   |        | | | |:  |:
Db         21 FEGEKVFRVNVEDENHINLLHTLASTTQIDFWKPDSVTQIKPHSTADFRVKAEDILTVED 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||: ::  ::: |         |:   ||:|:    | :| : :|  :||
Db         81 FLKQNELHYEVLINNLRLVLEGQFGRQV---PATGHSYEKYNRWETIEAWTQQVTSENPD 137

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::  ||::||   :|:||| ||     | ||::||| ||||||||||  ||:
Db        138 LISRRSIGTTFEGRTIYLLKV-GKAGQNKPAIFMDCGFHAREWISPAFWQWFVREXIRTY 196

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      |||||| ||     :|:|||  |||
Db        197 GQEIHMTELLDKLDFYVLPVGNIDGYVYTWTKNRMWRKTRSTQVGTNCVGTDPTRNF-DA 255

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||: ||| : | |||||   ||| | ||:|:|:| |:: ||||:::|||| :::|||
Db        256 GWCKIGASRNPCDETYCGPAAESEKETKALANFIRSNLSSIKAYLTIHSYSQMMLYPYSY 315

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              :::: ||: :|   |: : |   |:||:| |: |:| | || ||| || ||||||
Db        316 DYKLTENNAELNALAKATVKEL-ATLHGTKYTYGPGATTIYPAAGGSDDWAYDQGIKYSF 374

Qy        324 T 324
              |
Db        375 T 375


RESULT 6
CBPC_HUMAN
ID   CBPC_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15088; Q96E94;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=90083291; PubMed=2594780;
RA   Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA   Austen K.F., Serafin W.E.;
RT   "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT   comparison of the protein with mouse mast cell carboxypeptidase A and
RT   rat pancreatic carboxypeptidases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mast cells;
RX   MEDLINE=92105393; PubMed=1729276;
RA   Reynolds D.S., Gurley D.S., Austen K.F.;
RT   "Cloning and characterization of the novel gene for mast cell
RT   carboxypeptidase A.";
RL   J. Clin. Invest. 89:273-282(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
```

```
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 110-417 FROM N.A.
RX   MEDLINE=92333165; PubMed=1629626;
RA   Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,
RA   Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT   "Human skin mast cell carboxypeptidase: functional characterization,
RT   cDNA cloning, and genealogy.";
RL   J. Invest. Dermatol. 99:138-145(1992).
RN   [5]
RP   SEQUENCE OF 110-137.
RX   MEDLINE=89214692; PubMed=2708524;
RA   Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT   "Human mast cell carboxypeptidase. Purification and
RT   characterization.";
RL   J. Clin. Invest. 83:1630-1636(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27717; AAA35652.1; -.
DR   EMBL; M73720; AAA59568.1; -.
DR   EMBL; M73716; AAA59568.1; JOINED.
DR   EMBL; M73717; AAA59568.1; JOINED.
DR   EMBL; M73718; AAA59568.1; JOINED.
DR   EMBL; M73719; AAA59568.1; JOINED.
DR   EMBL; BC012613; AAH12613.1; -.
DR   EMBL; S40234; AAB22578.2; ALT_SEQ.
DR   PIR; A43929; A43929.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   Genew; HGNC:2298; CPA3.
DR   MIM; 114851; -.
DR   GO; GO:0030141; C:secretory granule; NAS.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    15
FT   PROPEP      16   109       ACTIVATION PEPTIDE.
FT   CHAIN      110   417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL      176   176       ZINC (BY SIMILARITY).
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173   186       BY SIMILARITY.
FT   DISULFID   245   268       BY SIMILARITY.
FT   CONFLICT    63    63       N -> K (IN REF. 3).
SQ   SEQUENCE   417 AA;  48700 MW;  8CB90DB758117B24 CRC64;
```

```
  Query Match           34.6%;  Score 622;  DB 1;  Length 417;
  Best Local Similarity  35.5%;  Pred. No. 5.1e-46;
  Matches  128;  Conservative   71;  Mismatches  118;  Indels   44;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |  :|    |: :|  ::::|  | |:  | |     :     | | |:  :   :::
Db         21 FDREKVFRVKPQDEKQADIIKDLAKTNELDFWYPGATHHVAANMMVDFRVSEKESQAIQS 80

Qy         61 HLNVSGIPCSVLLADVEDLIQQQIS-NDTVSPRASASYYEQYHSLNEIYSWIEFITERHP 119
               |: : :   :|: |::: |::|    : :  | |   | :|::  :| :| | : :::|
Db         81 ALDQNKMHYEILIHDLQEEIEKQFDVKEDIPGRHS---YAKYNNWEKIVAWTEKMMDKYP 137

Qy        120 DMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------ 173
              :|:::| |||: |  ||||||: |::   : ||::|||||||||:|||| ||:
Db        138 EMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCGIHAREWVSPAFCQWFVYQATKT 196

Qy        174 -------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFAS 202
                                        |:      |||||||||   |: ||||||||| :
Db        197 YGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWRKNRSKNQNSKCIGTDLNRNF-N 255

Qy        203 KHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYS 262
                |     ::  |:: | |  |||| | ||| :|:| ::|:|| ||: ||||| ::|||
Db        256 ASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSHLNEIKVYITFHSYSQMLLFPYG 315

Qy        263 YTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYS 322
              ||     :||:|: ||       :  |   ||| :|    |:|   |   || |||||::
Db        316 YTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIESTIYPISGSSLDWAYDLGIKHT 374

Qy        323 F 323
              |
Db        375 F 375
```

```
RESULT 7
CBPB_BOVIN
ID   CBPB_BOVIN     STANDARD;      PRT;   306 AA.
AC   P00732;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B (EC 3.4.17.2).
GN   CPB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=75217824; PubMed=1057162;
RA   Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT   "Amino-acid sequence of bovine carboxypeptidase B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN   [2]
RP   SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX   MEDLINE=74260705; PubMed=4833744;
RA   Schmidt J.J., Hirs C.H.W.;
RT   "Primary structure of bovine carboxypeptidase B. Inferences from the
RT   locations of the half-cystines and identification of the active site
RT   arginine.";
RL   J. Biol. Chem. 249:3756-3764(1974).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX   MEDLINE=76265065; PubMed=957425;
RA   Schmid M.F., Herriott J.R.;
RT   "Structure of carboxypeptidase B at 2.8-A resolution.";
RL   J. Mol. Biol. 103:175-190(1976).
RN   [4]
RP   ACTIVE SITE.
RX   MEDLINE=70007159; PubMed=5344132;
RA   Plummer T.H. Jr.;
RT   "Isolation and sequence of peptides at the active center of bovine
RT   carboxypeptidase B.";
```

```
RL   J. Biol. Chem. 244:5246-5253(1969).
RN   [5]
RP   ACTIVE SITE.
RX   MEDLINE=73061487; PubMed=4565668;
RA   Kimmel M.T., Plummer T.H. Jr.;
RT   "Identification of a glutamic acid at the active center of bovine
RT   carboxypeptidase B.";
RL   J. Biol. Chem. 247:7864-7869(1972).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A93797; CPBOB.
DR   PDB; 1CPB; 30-SEP-83.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT   DISULFID     63     76
FT   DISULFID    135    158
FT   DISULFID    149    163
FT   METAL        66     66       ZINC.
FT   METAL        69     69       ZINC.
FT   METAL       194    194       ZINC.
FT   ACT_SITE    246    246
FT   ACT_SITE    268    268       NUCLEOPHILE.
SQ   SEQUENCE   306 AA;  34612 MW;  C329D2655C44A172 CRC64;

  Query Match             32.6%;  Score 586;  DB 1;  Length 306;
  Best Local Similarity   43.9%;  Pred. No. 4.2e-43;
  Matches  116;  Conservative   44;  Mismatches   64;  Indels   40;  Gaps    5;

Qy         98 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 157
              ||:|::   | :| | :   :||:::: ||::|    :|:||| ||  : | |:::|||
Db          6 YEKYNNWETIEAWTEQVASENPDLISRSAIGTTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy        158 IHAREWISPAFCLWFI------------------------------GH------NRMWR 180
              |||||||||||| ||:                              |:      |||||
Db         65 FHAREWISPAFCQWFVREAVRTYGREIHMTEFLDKLDFYVLPVVNIDGYIYTWTTNRMWR 124

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              | ||  | : | ||||||||    ||  |||:: |||||||   ||| | |||| |:| :
Db        125 KTRSTRAGSSCTGTDLNRNF-DAGWCSIGASNNPCSETYCGSAAESEKESKAVADFIRNH 183

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              :: ||||:::||||| :::|||||     |:: ||: :|  ||: :  :   | |::|. |:
Db        184 LSSIKAYLTIHSYSQMMLYPYSYDYKLPKNNVELNTLAKGAVKKL-ASLHGTTYSYGPGA 242

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFT 324
              |:| | || ||| || |||||||
Db        243 TTIYPASGGSDDWAYDQGIKYSFT 266


RESULT 8
CBPC_RAT
ID   CBPC_RAT       STANDARD;      PRT;   309 AA.
AC   P21961;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE   A3).
GN   CPA3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91105153; PubMed=1988052;
RA   Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA   Neurath H.;
RT   "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT   enzyme activity within mast cell granules.";
RL   Biochemistry 30:648-655(1991).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A38395; A38395.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   METAL        68     68       ZINC (BY SIMILARITY).
FT   METAL        71     71       ZINC (BY SIMILARITY).
FT   METAL       196    196       ZINC (BY SIMILARITY).
FT   ACT_SITE    248    248       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    270    270       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     65     78       BY SIMILARITY.
FT   DISULFID    137    160       BY SIMILARITY.
FT   VARIANT       1      1       MISSING (IN THE MAJOR FORM).
SQ   SEQUENCE   309 AA;  35786 MW;  20330FABC3EE83EF CRC64;

  Query Match             30.7%;  Score 552;  DB 1;  Length 309;
  Best Local Similarity   42.6%;  Pred. No. 3.6e-40;
  Matches  112;  Conservative   39;  Mismatches   72;  Indels   40;  Gaps    5;

Qy         98 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 157
              | :|:  |:| || | : |:||:|:::| |||: |  ||||||: |::   : ||::|||
Db          8 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 66

Qy        158 IHAREWISPAFCLWFI-------GHN-----------------------------RMWR 180
              ||||||:||||| ||:       | |                             ||||
Db         67 IHAREWVSPAFCQWFVYQAAKSYGKNNIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 126

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||   :: ||||||||||     |      : : |   | |  |||| | ||| :|:| :
Db        127 KNRSKNPSSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 185

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              :| |||||: ||||| ::||| ||      :|::|  ||  |   :  :   ||| :|  :
Db        186 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy        301 ETLYLAPGGGDDWIYDLGIKYSF 323
              |:|   |   || ||||||::|
Db        245 STIYKTSGSSLDWAYDLGIKHTF 267


RESULT 9
CBP2_RAT
ID   CBP2_RAT       STANDARD;      PRT;   417 AA.
AC   P19222;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=89034323; PubMed=3182871;
RA   Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA   Graf M., Rutter W.J.;
RT   "A novel rat carboxypeptidase, CPA2: characterization, molecular
```

```
RT   cloning, and evolutionary implications on substrate specificity in
RT   the carboxypeptidase gene family.";
RL   J. Biol. Chem. 263:17828-17836(1988).
RN   [2]
RP   SEQUENCE OF 131-143 FROM N.A.
RX   MEDLINE=95386501; PubMed=7657630;
RA   Normant E., Gros C., Schwartz J.C.;
RT   "Carboxypeptidase A isoforms produced by distinct genes or
RT   alternative splicing in brain and other extrapancreatic tissues.";
RL   J. Biol. Chem. 270:20543-20549(1995).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=92105124; PubMed=1761558;
RA   Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT   "Structural evolution of an enzyme specificity. The structure of rat
RT   carboxypeptidase A2 at 1.9-A resolution.";
RL   J. Biol. Chem. 266:24606-24612(1991).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23721; AAA40956.1; -.
DR   EMBL; M23714; AAA40956.1; JOINED.
DR   EMBL; M23715; AAA40956.1; JOINED.
DR   EMBL; M23716; AAA40956.1; JOINED.
DR   EMBL; M23717; AAA40956.1; JOINED.
DR   EMBL; M23718; AAA40956.1; JOINED.
DR   EMBL; M23719; AAA40956.1; JOINED.
DR   EMBL; M23720; AAA40956.1; JOINED.
DR   EMBL; S79837; -; NOT_ANNOTATED_CDS.
DR   PIR; A32128; A32128.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.002; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17    112       ACTIVATION PEPTIDE.
FT   CHAIN       113    417       CARBOXYPEPTIDASE A2.
FT   METAL       177    177       ZINC.
FT   METAL       180    180       ZINC.
FT   METAL       304    304       ZINC.
FT   ACT_SITE    378    378       NUCLEOPHILE.
FT   DISULFID    246    269       BY SIMILARITY.
FT   DISULFID    318    352       PROBABLE.
SQ   SEQUENCE   417 AA;  46912 MW;  BEDCC41A830F2D45 CRC64;

  Query Match             30.6%;  Score 549.5;  DB 1;  Length 417;
  Best Local Similarity   34.6%;  Pred. No. 8.9e-40;
  Matches  134;  Conservative   63;  Mismatches  127;  Indels   63;  Gaps   12;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTT--YEIVLWQPVTADLIVKKKQVHFFVNASDVDNV 58
              |   |||  :|    |:: |  |       |:  |:  |    :  : ||  |  : :  |
Db         20 FVGDQVLEIIPSHEEQIRTLLQLEAEEHLELDFWKSPT----IPGETVHVRVPFASIQAV 75

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISN---DTVSPRASASYYEQYHSLNEIYSWIEFIT 115
              |  |   ||  |::: ||: |: |:      :    |      :| ||:| |||   :: :
Db         76 KVFLESQGIDYSIMIEDVQVLLDQEREEMLFNQQRERGGNFNFEAYHTLEEIYQEMDNLV 135

Qy        116 ERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIHAREWISPAFCLW--- 171
                :| :::|:::|||||  |: ||| | | :   | |||:| |||||||:: |  ||
Db        136 AENPGLVSKVNLGSSFENRPMNVLKFSTGGD---KPAIWLDAGIHAREWVTQATALWTAN 192

Qy        172 --------------------FI-------GH------NRMWRKNRSFYANNHCIGTDLN 197
                                 |:       |:      ||||| ||  : : |:| | |
Db        193 KIASDYGTDPAITSLLNTLDIFLLPVTNPDGYVFSQTTNRMWRKTRSKRSGSGCVGVDPN 252

Qy        198 RNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHI 257
              ||:   ::   ||||| ||::| |  | || |||::  |:: :  ::||:|::||||| :
Db        253 RNW-DANFGGPGASSSPCSDSYHGPKPNSEVEVKSIVDFIKSH-GKVKAFITLHSYSQLL 310

Qy        258 VFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDL 317
              :||| |  :|  |  ||  || :| :|: |    | |  |     :| | ||  || |||
Db        311 MFPYGYKCTKPDDFNELDEVAQKAAQAL-KRLHGTSYKVGPICSVIYQASGGSIDWAYDL 369

Qy        318 GIKYSFTSN----------PPVEKLLP 334
              ||||||             | :::||
Db        370 GIKYSFAFELRDTAFYGFLLPAKQILP 396


RESULT 10
CBPA_BOVIN
ID   CBPA_BOVIN     STANDARD;      PRT;   419 AA.
AC   P00730;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A precursor (EC 3.4.17.1).
GN   CPA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91151335; PubMed=1998496;
RA   le Hueerou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
RT   "Cloning and nucleotide sequence of a bovine pancreatic
RT   preprocarboxypeptidase A cDNA.";
RL   Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=96096549; PubMed=8522204;
RA   Goo J.H., Kim K.H., Choi K.Y.;
RT   "Cloning, sequencing and expression of the gene encoding a major
RT   allotypic preprocarboxypeptidase A from bovine pancreas.";
RL   Gene 165:333-334(1995).
RN   [3]
RP   SEQUENCE OF 111-417.
RX   MEDLINE=71155183; PubMed=5102489;
RA   Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT   chymotryptic peptides of the cyanogen bromide fragment F-I.";
RL   Biochemistry 10:938-950(1971).
RN   [4]
RP   REVISIONS TO 138 AND 141.
RX   MEDLINE=72138789; PubMed=5143102;
RA   Petra P.H., Hermodson M.A., Walsh K.A., Neurath H.;
RT   "Characterization of bovine carboxypeptidase A (Allan).";
RL   Biochemistry 10:4023-4025(1971).
RN   [5]
RP   SEQUENCE OF 17-120.
RX   MEDLINE=89150306; PubMed=3147705;
RA   Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neurath H.;
RT   "The amino acid sequence of the activation peptide of bovine pro-
RT   carboxypeptidase A.";
RL   Biochimie 70:1137-1142(1988).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
```

```
RX   MEDLINE=83294519; PubMed=6887246;
RA   Rees D.C., Lewis M., Lipscomb W.N.;
RT   "Refined crystal structure of carboxypeptidase A at 1.54-A
RT   resolution.";
RL   J. Mol. Biol. 168:367-387(1983).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX   MEDLINE=96003618; PubMed=7556081;
RA   Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT   "The three-dimensional structure of the native ternary complex of
RT   bovine pancreatic procarboxypeptidase A with proproteinase E and
RT   chymotrypsinogen C.";
RL   EMBO J. 14:4387-4394(1995).
RN   [8]
RP   VARIANT ALLELIC.
RX   MEDLINE=69283620; PubMed=5817619;
RA   Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Identification of the amino acid replacements characterizing the
RT   allotypic forms of bovine carboxypeptidase A.";
RL   Biochemistry 8:2762-2768(1969).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC       COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC       PROPROTEINASE E.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61851; AAA30426.1; -.
DR   EMBL; M61851; AAA30427.1; -.
DR   EMBL; Z33906; CAA83955.1; -.
DR   PIR; JN0126; CPBOA.
DR   PDB; 3CPA; 15-JAN-87.
DR   PDB; 4CPA; 22-OCT-84.
DR   PDB; 5CPA; 15-JAN-87.
DR   PDB; 6CPA; 15-OCT-91.
DR   PDB; 7CPA; 31-JAN-94.
DR   PDB; 8CPA; 31-JAN-94.
DR   PDB; 1CBX; 31-JAN-94.
DR   PDB; 1CPS; 15-OCT-94.
DR   PDB; 2CTB; 31-JAN-94.
DR   PDB; 2CTC; 31-JAN-94.
DR   PDB; 1ARL; 01-AUG-96.
DR   PDB; 1ARM; 17-AUG-96.
DR   PDB; 1BAV; 01-APR-97.
DR   PDB; 1YME; 12-FEB-97.
DR   PDB; 1CPX; 05-AUG-98.
DR   PDB; 1PYT; 27-JAN-97.
DR   PDB; 1ELL; 26-JUN-02.
DR   PDB; 1ELM; 26-JUN-02.
DR   PDB; 1F57; 06-SEP-00.
DR   PDB; 1HDQ; 15-NOV-01.
DR   PDB; 1HDU; 15-NOV-01.
DR   PDB; 1HEE; 23-NOV-01.
DR   PDB; 1M4L; 04-MAR-03.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure; Polymorphism.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A.
FT   METAL      179   179       ZINC.
FT   METAL      182   182       ZINC.
FT   METAL      306   306       ZINC.
FT   ACT_SITE   358   358       PROTON DONOR.
FT   ACT_SITE   380   380       NUCLEOPHILE.
FT   DISULFID   248   271
FT   VARIANT    289   289       I -> V (IN ALLELIC VARIANT).
FT   VARIANT    338   338       E -> A (IN ALLELIC VARIANT).
FT   VARIANT    415   415       L -> V (IN ALLELIC VARIANT).
FT   CONFLICT    95    95       S -> L (IN REF. 5).
FT   CONFLICT   199   199       D -> N (IN REF. 3).
FT   CONFLICT   203   203       D -> N (IN REF. 3).
FT   CONFLICT   224   224       D -> N (IN REF. 3).
FT   CONFLICT   232   232       Q -> E (IN REF. 3).
FT   CONFLICT   295   295       D -> N (IN REF. 3).
FT   TURN       114   116
FT   TURN       119   120
FT   HELIX      125   138
FT   TURN       140   142
FT   STRAND     143   150
FT   TURN       152   153
FT   STRAND     156   162
FT   STRAND     171   176
FT   TURN       180   181
FT   HELIX      183   199
FT   TURN       200   202
FT   HELIX      204   212
FT   STRAND     214   218
FT   HELIX      223   231
FT   TURN       232   232
FT   TURN       234   235
FT   STRAND     242   242
FT   TURN       244   245
FT   STRAND     249   249
FT   TURN       253   254
FT   TURN       261   262
FT   STRAND     267   267
FT   TURN       270   271
FT   TURN       273   274
FT   STRAND     275   275
FT   TURN       280   281
FT   HELIX      284   296
FT   STRAND     299   307
FT   STRAND     311   314
FT   TURN       324   325
FT   HELIX      326   344
FT   STRAND     349   352
FT   HELIX      353   356
FT   TURN       357   357
FT   HELIX      364   370
FT   TURN       371   372
FT   STRAND     375   381
FT   TURN       388   389
FT   HELIX      393   395
FT   HELIX      396   416
SQ   SEQUENCE   419 AA;  47082 MW;  21B86407B3BFC452 CRC64;

  Query Match             29.8%;  Score 536.5;  DB 1;  Length 419;
  Best Local Similarity   34.8%;  Pred. No. 1.2e-38;
  Matches  129;  Conservative   59;  Mismatches  124;  Indels   59;  Gaps   10;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--W----QPVTADLIVKKKQVHFFVNASD 54
              |   |||         :|| :: |      : |  |     ||           :   |
Db         20 FVGHQVLRITAADEAEVQTVKELEDLEHLQLDFWRGPGQP--------GSPIDVRVPFPS 71

Qy         55 VDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSW 110
              :  ||  |  ||   ::: ||: |:   |:|:       |:: :: |  ||:|:||| :
Db         72 LQAVKVFLEAHGIRYRIMIEDVQSLLDEEQEQMFASQSRARSTNTFNYATYHTLDEIYDF 131
```

```
Qy        111 IEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCL 170
              :: :   || :::|: || |:|  |:|||| |    : : ||||| |||:||||: |  :
Db        132 MDLLVAEHPQLVSKLQIGRSYEGRPIYVLKFS-TGGSNRPAIWIDLGIHSREWITQATGV 190

Qy        171 WFI-----------------------------------GHNRMWRKNRSFYANNHCIG 193
              ||                                  ||:||| ||  ::: |:|
Db        191 WFAKKFTEDYGQDPSFTAILDSMDIFLEIVTNPDGFAFTHSQNRLWRKTRSVTSSSLCVG 250

Qy        194 TDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSY 253
              | |||:    : : ||||| ||||| | |  || |||::  |:: : |  ||::|:|||
Db        251 VDANRNW-DAGFGKAGASSSPCSETYHGKYANSEVEVKSIVDFVKDHGN-FKAFLSIHSY 308

Qy        254 SQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDW 313
              || :::|| ||     |  ||: ||  || |: |:   | | :|    |:| | ||  ||
Db        309 SQLLLYPYGYTTQSIPDKTELNQVAKSAVEAL-KSLYGTSYKYGSIITTIYQASGGSIDW 367

Qy        314 IYDLGIKYSFT 324
              |: |||||||
Db        368 SYNQGIKYSFT 378


RESULT 11
CBP1_RAT
ID   CBP1_RAT       STANDARD;      PRT;   419 AA.
AC   P00731;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82105986; PubMed=6275388;
RA   Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA   Pictet R.L., Valenzuela P., Rutter W.J.;
RT   "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT   characterization of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89034324; PubMed=3182872;
RA   Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT   "Structural characterization of the rat carboxypeptidase A1 and B
RT   genes. Comparative analysis of the rat carboxypeptidase gene
RT   family.";
RL   J. Biol. Chem. 263:17837-17845(1988).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01232; CAA24542.1; -.
DR   EMBL; J00713; AAA40893.1; -.
DR   EMBL; M23990; AAA40955.1; -.
DR   EMBL; M23960; AAA40955.1; JOINED.
DR   EMBL; M23985; AAA40955.1; JOINED.
DR   EMBL; M23986; AAA40955.1; JOINED.
DR   EMBL; M23987; AAA40955.1; JOINED.
DR   EMBL; M23988; AAA40955.1; JOINED.
DR   EMBL; M23989; AAA40955.1; JOINED.
DR   PIR; A00911; CPRTA.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A1.
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      182   182       ZINC (BY SIMILARITY).
FT   METAL      306   306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358   358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380   380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248   271       BY SIMILARITY.
FT   CONFLICT   196   196       I -> V (IN REF. 1).
FT   CONFLICT   261   263       FGM -> LGK (IN REF. 1).
FT   CONFLICT   347   347       K -> E (IN REF. 1).
SQ   SEQUENCE   419 AA;  47197 MW;  BB002D1CB99B7491 CRC64;

  Query Match             29.7%;  Score 533.5;  DB 1;  Length 419;
  Best Local Similarity   35.4%;  Pred. No. 2.1e-38;
  Matches  130;  Conservative   59;  Mismatches  127;  Indels   51;  Gaps    9;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTADLIVKKKQVHFFVNASDVDNV 58
              |   |||         ||| :: |     : |  |:       |    :| |       : :|
Db         20 FVGHQVLRISAADEAQVQKVKELEDLEHLQLDFWRDAARAGIPIDVRVPF----PSIQSV 75

Qy         59 KAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFI 114
              || |   ||   ::: ||: |:   :||:|       :: |: |  ||:|:||| ::: :
Db         76 KAFLEYHGISYEIMIEDVQLLLDEEKQQMSAFQARALSTDSFNYATYHTLDEIYEFMDLL 135

Qy        115 TERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIG 174
              || :::|| ||::||  |::||| |    | : ||||| |||:|||:: |  :||
Db        136 VAEHPQLVSKIQIGNTFEGRPIHVLKFS-TGGTNRPAIWIDTGIHSREWVTQASGVWFAK 194

Qy        175 H-----------------------------------NRMWRKNRSFYANNHCIGTDLN 197
                                                  ||||| ||    : |:| | |
Db        195 KITKDYGQDPTFTAVLDNMDIFLEIVTNPDGFAYTHKTNRMWRKTRSHTQGSLCVGVDPN 254

Qy        198 RNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHI 257
              ||:    :   ||||: ||||| | :| || |||::  |:  : | |||:||:|||| :
Db        255 RNW-DAGFGMAGASSNPCSETYRGKFPNSEVEVKSIVDFVTSHGN-IKAFISIHSYSQLL 312

Qy        258 VFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDL 317
              ::|| ||   : |  ||  :|  || |:  :   |:: :|   :|:| | |   || |
Db        313 LYPYGYTSEPAPDQAELDQLAKSAVTAL-TSLHGTKFKYGSIIDTIYQASGSTIDWTYSQ 371

Qy        318 GIKYSFT 324
              |||||||
Db        372 GIKYSFT 378


RESULT 12
CPB2_HUMAN
ID   CPB2_HUMAN     STANDARD;      PRT;   417 AA.
AC   P48052; Q96A12; Q96QN3;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
```

```
RC   TISSUE=Pancreas;
RX   MEDLINE=95204457; PubMed=7896805;
RA   Catasus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
RA   Billeter M.;
RT   "The sequence and conformation of human pancreatic
RT   procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT   three-dimensional model.";
RL   J. Biol. Chem. 270:6651-6657(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98111000; PubMed=9450539;
RA   Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,
RA   Aviles F.X.;
RT   "Characterisation and preliminary X-ray diffraction analysis of human
RT   pancreatic procarboxypeptidase A2.";
RL   FEBS Lett. 420:7-10(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98046021; PubMed=9384570;
RA   Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT   "The three-dimensional structure of human procarboxypeptidase A2.
RT   Deciphering the basis of the inhibition, activation and intrinsic
RT   activity of the zymogen.";
RL   EMBO J. 16:6906-6913(1997).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19977; AAA74425.1; -.
DR   EMBL; BC007009; AAH07009.1; -.
DR   EMBL; BC014571; AAH14571.1; -.
DR   EMBL; BC015140; AAH15140.1; -.
DR   PIR; A56171; A56171.
DR   PDB; 1AYE; 13-JAN-99.
DR   PDB; 1DTD; 12-JUL-00.
DR   PDB; 1O6X; 30-JAN-03.
DR   MEROPS; M14.002; -.
DR   Genew; HGNC:2297; CPA2.
DR   MIM; 600688; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0007039; P:vacuolar protein catabolism; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   112       ACTIVATION PEPTIDE.
FT   CHAIN      113   417       CARBOXYPEPTIDASE A2.
FT   METAL      177   177       ZINC.
FT   METAL      180   180       ZINC.
FT   METAL      304   304       ZINC.
FT   ACT_SITE   378   378       NUCLEOPHILE.
FT   DISULFID   246   269
FT   DISULFID   318   352
FT   CONFLICT    80    80       E -> G (IN REF. 2; AAH07009).
FT   CONFLICT   302   302       T -> I (IN REF. 1).
SQ   SEQUENCE   417 AA;  46828 MW;  3D9C073A98700A83 CRC64;

  Query Match             28.6%;  Score 514.5;  DB 1;  Length 417;
  Best Local Similarity   33.1%;  Pred. No. 9.2e-37;
  Matches  128;  Conservative   62;  Mismatches  134;  Indels   63;  Gaps   10;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTADLIVKKKQVHFFVNASDVDNV 58
              |   |||  :|    |:: |  |     : |  |:  |       :  |  |   :|  |
Db         20 FVGDQVLEIVPSNEEQIKNLLQLEAQEHLQLDFWKSPT----TPGETAHVRVPFVNVQAV 75

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS---YYEQYHSLNEIYSWIEFIT 115
              |  |    ||  |::: ||: |: ::      : |   |    :  ||:| ||   :: :
Db         76 KVFLESQGIAYSIMIEDVQVLLDKENEEMLFNRRRERSGNFNFGAYHTLEEISQEMDNLV 135

Qy        116 ERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIHAREWISPAFCLWFIG 174
              || :::|::|||||  |: ||| | | :   | |||:| |||||||:: |  ||
Db        136 AEHPGLVSKVNIGSSFENRPMNVLKFSTGGD---KPAIWLDAGIHAREWVTQATALWTAN 192

Qy        175 ----------------------------------HNRMWRKNRSFYANNHCIGTDLN 197
                                                 |||||| ||  : : |:| | |
Db        193 KIVSDYGKDPSITSILDALDIFLLPVTNPDGYVFSQTKNRMWRKTRSKVSGSLCVGVDPN 252

Qy        198 RNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHI 257
              ||:   :   ||||: ||::| |    || |||::  |:: :  ::||:|::||||| :
Db        253 RNW-DAGFGGPGASSNPCSDSYHGPSANSEVEVKSIVDFIKSH-GKVKAFITLHSYSQLL 310

Qy        258 VFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDL 317
              :||| |  :|  | :||| || :| ::: ::   |:|  |     :| | ||  || ||
Db        311 MFPYGYKCTKLDDFDELSEVAQKAAQSL-RSLHGTKYKVGPICSVIYQASGGSIDWSYDY 369

Qy        318 GIKYSFTSN----------PPVEKLLP 334
              ||||||               |  ::||
Db        370 GIKYSFAFELRDTGRYGFLLPARQILP 396


RESULT 13
CBP1_HUMAN
ID   CBP1_HUMAN     STANDARD;      PRT;   419 AA.
AC   P15085; Q9BS67;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1 OR CPA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
```

```
RX   MEDLINE=93038569; PubMed=1417781;
RA   Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA   Wicker-Planquart C., Puigserver A.;
RT   "cDNA cloning and sequence analysis of human pancreatic
RT   procarboxypeptidase A1.";
RL   Biochem. J. 287:299-303(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 17-42.
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67318; CAA47732.1; -.
DR   EMBL; BC005279; AAH05279.1; -.
DR   PIR; S29127; S29127.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   Genew; HGNC:2296; CPA1.
DR   MIM; 114850; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1     16
FT   PROPEP      17    110       ACTIVATION PEPTIDE.
FT   CHAIN      111    419       CARBOXYPEPTIDASE A1.
FT   METAL      179    179       ZINC (BY SIMILARITY).
FT   METAL      182    182       ZINC (BY SIMILARITY).
FT   METAL      306    306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358    358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380    380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248    271       BY SIMILARITY.
FT   CONFLICT   276    276       H -> R (IN REF. 2).
SQ   SEQUENCE   419 AA;  47140 MW;  439FAFFFAEE958B1 CRC64;

  Query Match             27.4%;  Score 493.5;  DB 1;  Length 419;
  Best Local Similarity   33.1%;  Pred. No. 5.9e-35;
  Matches  126;  Conservative   58;  Mismatches  118;  Indels   79;  Gaps   12;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQ-------PVTADLIVKKKQVHFFVN 51
              |   |||         ||| :: |     : |  |:       |:         :| |
Db         20 FVGHQVLRISVADEAQVQKVKELEDLEHLQLDFWRGPAHPGSPIDV-------RVPF--- 69

Qy         52 ASDVDNVKAHLNVSGIPCSVLLADVEDLI---QQQI--------SNDTVSPRASASYYEQ 100
                :  ||  |   ||    :: ||: |:   |:|:         | || :       |
Db         70 -PSIQAVKIFLESHGISYETMIEDVQSLLDEEQEQMFAFRSRARSTDTFN-------YAT 121

Qy        101 YHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHA 160
              ||:| ||| ::: :   :| :::|| ||:::|  |:|||| |    : : ||||| |||:
Db        122 YHTLEEIYDFLDLLVAENPHLVSKIQIGNTYEGRPIYVLKFS-TGGSKRPAIWIDTGIHS 180

Qy        161 REWISPAFCLWFI-----------------------------------GHNRMWRKNR 183
              |||:: |  :||                                   ||||| |
Db        181 REWVTQASGVWFAKKITQDYGQDAAFTAILDTLDIFLEIVTNPDGFAFTHSTNRMWRKTR 240

Qy        184 SFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQ 243
              |  | : ||| | |||: :      ||||: ||||| | :  || |||::  |:: : |
Db        241 SHTAGSLCIGVDPNRNWDAGFGL-SGASSNPCSETYHGKFANSEVEVKSIVDFVKDHGN- 298

Qy        244 IKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETL 303
              |||:||:||||| :::|| |      | :||  ::  || |:  :   |:: :|   : :
Db        299 IKAFISIHSYSQLLMYPYGYKTEPVPDQDELDQLSKAAVTAL-ASLYGTKFNYGSIIKAI 357

Qy        304 YLAPGGGDDWIYDLGIKYSFT 324
              | | |   || |  |||||||
Db        358 YQASGSTIDWTYSQGIKYSFT 378


RESULT 14
CBP4_HUMAN
ID   CBP4_HUMAN     STANDARD;      PRT;   421 AA.
AC   Q9UI42;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN   CPA4 OR CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99310537; PubMed=10383164;
RA   Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT   "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT   deacetylase inhibitors during differentiation of prostate epithelial
RT   cancer cells.";
RL   Cancer Res. 59:2981-2988(1999).
RN   [2]
RP   IMPRINTING.
RX   MEDLINE=22483629; PubMed=12552318;
RA   Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA   Yoshiura K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA   Niikawa N., Kishino T.;
RT   "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT   imprinting domain.";
RL   Hum. Genet. 112:220-226(2003).
CC   -!- FUNCTION: Could be involved in the histone hyperacetylation
CC       pathway.
CC   -!- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,
```

```
CC       heart, intestine, kidney, liver and lung. Except for fetal brain
CC       that shows no imprinting, expression was found preferentially from
CC       the maternal allele.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095719; AAF23230.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.017; -.
DR   Genew; HGNC:15740; CPA4.
DR   MIM; 607635; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; NAS.
DR   GO; GO:0016573; P:histone acetylation; NAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   113       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN      114   421       CARBOXYPEPTIDASE A4.
FT   METAL      181   181       ZINC (BY SIMILARITY).
FT   METAL      184   184       ZINC (BY SIMILARITY).
FT   METAL      308   308       ZINC (BY SIMILARITY).
FT   ACT_SITE   382   382       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   250   273       BY SIMILARITY.
SQ   SEQUENCE   421 AA;  47379 MW;  91883D25B705BC6A CRC64;

  Query Match             27.3%;  Score 491;  DB 1;  Length 421;
  Best Local Similarity   32.7%;  Pred. No. 9.8e-35;
  Matches  123;  Conservative   59;  Mismatches  126;  Indels   68;  Gaps   11;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVL--WQPVTA-----DLIVKKKQVHFFVNAS 53
              |   |||     |   ::  | |  :  : |  |:  ::      |::|    :   |
Db         21 FFGDQVLRINVRNGDEISKLSQLVNSNNLKLNFWKSPSSFNRPVDVLVPSVSLQAF---- 76

Qy         54 DVDNVKAHLNVSGIPCSVLLADVEDLI---QQQISNDTVSPRASASY-YEQYHSLNEIYS 109
                 |: |   |:  :| : |:: |:     :: ::    |:| :: |  ||||  ||
Db         77 -----KSFLRSQGLEYAVTIEDLQALLDNEDDEMQHNEGQERSSNNFNYGAYHSLEAIYH 131

Qy        110 WIEPITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFC 169
               :: |    ||:  :: || |||  |:|||| |  :   : |:|:: |||:|||| |
Db        132 EMDNIAADFPDLARRVKIGHSFENRPMYVLKFSTGKGVRRPAVWLNAGIHSREWISQATA 191

Qy        170 LW-----------------------FI------------GHNRMWRKNRSFYANNHCI 192
              :|                       |:              ||:||| ||    : ||
Db        192 IWTARKIVSDYQRDPAITSILEKMDIFLLPVANPDGYVYTQTQNRLWRKTRSRNPGSSCI 251

Qy        193 GTDLNRN----FASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYI 248
              | | |||    || |     ||| : ||| | | :  || |||:|  |:::: |  | :|
Db        252 GADPNRNWNASFAGK-----GASDNPCSEVYHGPHANSEVEVKSVVDFIQKHGN-FKGFI 305

Qy        249 SMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPG 308
              :||||| :::|| |:  |: | |||  ||  | :|:   |  | |  |     |:| | |
Db        306 DLHSYSQLLMYPYGYSVKKAPDAEELDKVARLAAKALASVS-GTEYQVGPTCTTVYPASG 364

Qy        309 GGDDWIYDLGIKYSFT 324
                || || |||::||
Db        365 SSIDWAYDNGIKFAFT 380
```

```
RESULT 15
CBPZ_SIMVI
ID   CBPZ_SIMVI     STANDARD;      PRT;   304 AA.
AC   P42788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS   Simulium vittatum (Black fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC   Simuliidae; Simulium.
OX   NCBI_TaxID=7192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Gut;
RX   MEDLINE=94093864; PubMed=8269093;
RA   Ramos A., Mahowald A., Jacobs-Lorena M.;
RT   "Gut-specific genes from the black fly Simulium vittatum encoding
RT   trypsin-like and carboxypeptidase-like proteins.";
RL   Insect Mol. Biol. 1:149-163(1993).
CC   -!- FUNCTION: Involved in the digestion of the blood meal.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: Gut-specific.
CC   -!- INDUCTION: By blood meal.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08481; AAA18531.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.UPA; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   NON_TER      1      1
FT   METAL       58     58       ZINC (BY SIMILARITY).
FT   METAL       61     61       ZINC (BY SIMILARITY).
FT   METAL      184    184       ZINC (BY SIMILARITY).
FT   ACT_SITE   236    236       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   259    259       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   125    148       BY SIMILARITY.
SQ   SEQUENCE   304 AA;  34849 MW;  2E6E3FF8A6AA9144 CRC64;

  Query Match             22.9%;  Score 411.5;  DB 1;  Length 304;
  Best Local Similarity   33.3%;  Pred. No. 4.4e-28;
  Matches   94;  Conservative   52;  Mismatches   85;  Indels   51;  Gaps    9;

Qy        100 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 159
              |||:| |||||:: : : ||: :  :  | |:|   :  :||| |:      : ::  ||
Db          1 QYHTLPEIYSWLDRLVQEHPEHVEPVVGGKSYEGREIRGVKVSYKK--GNPVVMVESNIH 58

Qy        160 AREWISPA---FCL-------------------WFI-------GH------NRMWRKNRS 184
              |||||: |   : |                   |:|       |:      :|||| ||
Db         59 AREWITAATTTYLLNELLTSKNSTIREMAENYDWYIFPVTNPDGYVYTHTTDRMWRKTRS 118

Qy        185 FYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQI 244
                :: | ||| |||: : || |:| ||  |:||| |    || | :: : ||:    ||
Db        119 PNPDSLCAGTDPNRNW-NFHWMEQGTSSRPCTETYGGKKAFSEVETRSFSDFLKTLKGQI 177

Qy        245 KAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLY 304
              | |:: ||||| ::||| :|   : :|::|  :   | |:: :    | || |:  : :|
Db        178 KVYLAFHSYSQLLLFPYGHTCQHTYNHDDLQAIGDAAARSLAQ-RYGTDYTVGNIYDAIY 236
```

```
Qy        305 LAPGGGDDWIY---DLGIKYSFTSNP---------PVEKLLP 334
              | ||  || |   |: | |::   |         |  :::|
Db        237 PASGGSMDWAYDTLDIPIAYTYELRPRDGWNGFQLPANQIIP 278
```

Search completed: January 1, 2004, 19:29:00
Job time : 15.4071 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:34:47 ; Search time 5219.6 Seconds
(without alignments)
2649.143 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135915_26829/app_query.fasta_1.1628
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_8325_@runat_31122003_135915_26829 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1784	99.2	1546	9	AB011969	AB011969 Homo sapi
2	1749.5	97.3	1625	6	AR232183	AR232183 Sequence
3	1732	96.3	1272	6	AR086324	AR086324 Sequence
4	1732	96.3	1272	6	BD106653	BD106653 Method of
5	1732	96.3	1728	6	AX409750	AX409750 Sequence
6	1732	96.3	1728	6	AX706330	AX706330 Sequence
7	1732	96.3	1728	9	HUMPCPBX	M75106 Human prepr
8	1732	96.3	1749	6	I16100	I16100 Sequence 2
9	1732	96.3	1749	6	I33526	I33526 Sequence 2
10	1727	96.1	1272	9	BT006936	BT006936 Homo sapi
11	1727	96.1	1272	12	BT007990	BT007990 Synthetic
12	1727	96.1	1715	9	BC007057	BC007057 Homo sapi
13	1494	83.1	1421	10	AB021968	AB021968 Mus muscu
14	1494	83.1	1490	10	AF164524	AF164524 Mus muscu
15	1484	82.5	1430	10	AF186188	AF186188 Mus muscu
16	1470	81.8	1425	10	AB042598	AB042598 Rattus no
17	697.5	38.8	1327	5	AF190274	AF190274 Bothrops
18	688	38.3	1441	5	AB099302	AB099302 Paralicht
19	646.5	36.0	1215	6	I67698	I67698 Sequence 1
20	646.5	36.0	1251	4	SSC133775	AJ133775 Sus scrof
21	637.5	35.5	1254	12	BT008111	BT008111 Synthetic
22	637.5	35.5	1263	6	A51896	A51896 Sequence 60
23	637.5	35.5	1263	6	A87530	A87530 Sequence 11
24	637.5	35.5	1263	6	AR085861	AR085861 Sequence
25	637.5	35.5	1284	6	A51913	A51913 Sequence 77
26	637.5	35.5	1284	6	AR085875	AR085875 Sequence
27	637.5	35.5	1310	9	HSA224866	AJ224866 Homo sapi
28	637.5	35.5	1454	9	BC015338	BC015338 Homo sapi
29	632	35.2	1442	10	MUSCARA	J05118 Mouse mast
30	630.5	35.1	1251	4	DOGZAP47	D78348 Canis famil
31	627	34.9	1427	10	RNU67914	U67914 Rattus norv
32	627	34.9	1674	9	BC012613	BC012613 Homo sapi
33	622	34.6	1622	9	HUMCARMC	M27717 Human mast
34	622	34.6	1622	11	G28614	G28614 human STS S
35	619.5	34.5	1302	6	AX616688	AX616688 Sequence
36	619.5	34.5	1311	6	AR225025	AR225025 Sequence
37	619.5	34.5	1311	6	AX106749	AX106749 Sequence
38	619.5	34.5	1993	6	AX704687	AX704687 Sequence
39	619.5	34.5	2154	6	A67368	A67368 Sequence 12
40	618	34.4	1332	6	AX333659	AX333659 Sequence
41	618	34.4	1332	9	HUMPASP	M81057 Human proca
42	617.5	34.3	1907	9	BC033684	BC033684 Homo sapi
43	610	33.9	927	6	AR072911	AR072911 Sequence
44	606.5	33.7	1672	5	BC041315	BC041315 Xenopus l
45	587	32.6	1339	5	GGCBPAMR	X64539 G.domesticu

ALIGNMENTS

RESULT 1

```
AB011969
LOCUS       AB011969                1546 bp    mRNA    linear   PRI 02-FEB-2000
DEFINITION  Homo sapiens mRNA for carboxypeptidase B-like protein, complete
            cds.
ACCESSION   AB011969
VERSION     AB011969.1  GI:6855463
KEYWORDS    carboxypeptidase B-like protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
REFERENCE   2  (sites)
  AUTHORS   Matsumoto,A.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1546)
  AUTHORS   Matsumoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
            Medicine, Department of Radiation Biophysics & Genetics;
            Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:amat@med.kobe-u.ac.jp, Tel:81-78341-7451)
COMMENT     Sequence updated (10-Apr-1998).
FEATURES             Location/Qualifiers
     source          1. .1546
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             20. .1102
                     /note="carboxypeptidase B-like protein in human brain"
                     /codon_start=1
                     /product="carboxypeptidase B-like protein"
                     /protein_id="BAA90475.1"
                     /db_xref="GI:6855464"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQTGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDL
                     NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYS
                     QHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDD
                     WIYDLGIKYSFTSNPPVEKLLPLSLK"
     polyA_site      1546
                     /note="21 A nucleotides"
BASE COUNT      453 a    323 c    297 g    473 t
ORIGIN

Alignment Scores:
Pred. No.:              2.47e-167       Length:         1546
Score:                  1784.00         Matches:        335
Percent Similarity:     99.41%          Conservative:   1
Best Local Similarity:  99.11%          Mismatches:     2
Query Match:            99.22%          Indels:         0
DB:                     9               Gaps:           0

US-09-980-881A-3 (1-338) x AB011969 (1-1546)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTCAGACTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 505

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        506 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATTCATGCC 565

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        566 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 625

Qy        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        626 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTT 685

Qy        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 GCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 745

Qy        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 805

Qy        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProPro 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 865

Qy        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 925

Qy        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        926 GCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 985

Qy        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1045

Qy        321 TyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 TATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1099

RESULT 2
AR232183
LOCUS       AR232183                1625 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6455294.
ACCESSION   AR232183
VERSION     AR232183.1  GI:27274083
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1625)
  AUTHORS   Gan,W., Ketchum,K.A., Di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human protease proteins, nucleic acid molecules encoding
            human protease proteins, and uses thereof
```

```
  JOURNAL   Patent: US 6455294-A 1 24-SEP-2002;
FEATURES             Location/Qualifiers
     source          1. .1625
                     /organism="unknown"
BASE COUNT      488 a    334 c    315 g    488 t
ORIGIN

Alignment Scores:
Pred. No.:              7.02e-164      Length:         1625
Score:                  1749.50        Matches:        336
Percent Similarity:     94.38%         Conservative:   0
Best Local Similarity:  94.38%         Mismatches:     2
Query Match:            97.30%         Indels:         18
DB:                     6              Gaps:           1

US-09-980-881A-3 (1-338) x AR232183 (1-1625)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 142

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 202

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 262

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 322

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 382

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 442

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 502

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        503 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 562

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 622

Qy        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTT 682

Qy        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 GCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 742

Qy        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 802

Qy        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 862

Qy        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 922

Qy        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
              |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        923 GCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 982

Qy        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1042

Qy        321 TyrSerPhe---------------------------------------------------- 323
              |||||||||
Db       1043 TATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGT 1102

Qy        324 -ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
               |||||||||||||||||||||||||||||||||||||||||||||
Db       1103 TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1148

RESULT 3
AR086324
LOCUS       AR086324               1272 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 1 from patent US 5985562.
ACCESSION   AR086324
VERSION     AR086324.1  GI:10013090
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.John. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk associated with plasma
            carboxypeptidase B polymorphisms
  JOURNAL   Patent: US 5985562-A 1 16-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="unknown"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN

Alignment Scores:
Pred. No.:              2.78e-162      Length:         1272
Score:                  1732.00        Matches:        338
Percent Similarity:     86.01%         Conservative:   0
Best Local Similarity:  86.01%         Mismatches:     0
Query Match:            96.33%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-3 (1-338) x AR086324 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 666

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||||
Db        667 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1207 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243
```

```
RESULT 4
BD106653
LOCUS       BD106653                1272 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Method of detecting thrombotic disease risk.
ACCESSION   BD106653
VERSION     BD106653.1  GI:23201471
KEYWORDS    JP 2002502253-A/1.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.J. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk
  JOURNAL   Patent: JP 2002502253-A 1 22-JAN-2002;
            SCHERING AG
COMMENT     PN   JP 2002502253-A/1
            PD   22-JAN-2002
            PF   02-JUN-1998 JP 1999501465
            PR   03-JUN-1997 US    08/869057
            PI   MICHAEL JOHN MORSER,MARIKO NAGASHIMA
            PC   C12Q1/68
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES                Location/Qualifiers
     source             1. .1272
                        /organism="synthetic construct"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32630"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                2.78e-162       Length:          1272
Score:                    1732.00         Matches:         338
Percent Similarity:       86.01%          Conservative:    0
Best Local Similarity:    86.01%          Mismatches:      0
Query Match:              96.33%          Indels:          55
DB:                       6               Gaps:            2

US-09-980-881A-3 (1-338) x BD106653 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 666

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||||
Db        667 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||
Db       1207 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 5
AX409750
LOCUS       AX409750               1728 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2397 from Patent WO0229103.
ACCESSION   AX409750
VERSION     AX409750.1  GI:21442455
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2397 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M75106"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:              4.16e-162       Length:         1728
Score:                  1732.00         Matches:        338
Percent Similarity:     86.01%          Conservative:   0
Best Local Similarity:  86.01%          Mismatches:     0
Query Match:            96.33%          Indels:         55
DB:                     6               Gaps:           2

US-09-980-881A-3 (1-338) x AX409750 (1-1728)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 505

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 565

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        566 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 625

Qy        175 ------------------------------------------------------------ 175

Db        626 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||||
Db        686 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 745

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 805

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 865

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 925

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 985

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1045

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1105

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1165

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1166 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1225

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||
Db       1226 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 6
```

```
AX706330
LOCUS       AX706330               1728 bp    DNA     linear   PAT 04-APR-2003
DEFINITION  Sequence 1 from Patent WO03014391.
ACCESSION   AX706330
VERSION     AX706330.1  GI:29562754
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Morten,J.E. and Jackson,K.N.
  TITLE     Method
  JOURNAL   Patent: WO 03014391-A 1 20-FEB-2003;
            AstraZeneca AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:                4.16e-162      Length:          1728
Score:                    1732.00        Matches:         338
Percent Similarity:       86.01%         Conservative:    0
Best Local Similarity:    86.01%         Mismatches:      0
Query Match:              96.33%         Indels:          55
DB:                       6              Gaps:            2

US-09-980-881A-3 (1-338) x AX706330 (1-1728)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 505

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 565

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||||
Db        566 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 625

Qy        175 ------------------------------------------------------------ 175

Db        626 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||
Db        686 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 745

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 805

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 865

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 925

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 985

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1045

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1105

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1165

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1166 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1225

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1226 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 7
HUMPCPBX
LOCUS       HUMPCPBX               1728 bp    mRNA    linear   PRI 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1  GI:189686
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1728)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
COMMENT     Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     gene            1. .1728
                     /gene="pCPB"
     CDS             20. .1291
                     /gene="pCPB"
                     /note="prepro-plasma carboxypeptidase B"
                     /codon_start=1
                     /protein_id="AAA60042.1"
                     /db_xref="GI:189687"
```

```
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
     mat_peptide     86. .1288
                     /gene="pCPB"
                     /product="plasma carboxypeptidase B"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:                4.16e-162      Length:          1728
Score:                    1732.00        Matches:         338
Percent Similarity:       86.01%         Conservative:    0
Best Local Similarity:    86.01%         Mismatches:      0
Query Match:              96.33%         Indels:          55
DB:                       9              Gaps:            2

US-09-980-881A-3 (1-338) x HUMPCPBX (1-1728)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 505

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 565

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||
Db        566 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 625

Qy        175 ------------------------------------------------------------ 175

Db        626 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 |||||||||||||||||||||||
Db        686 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 745

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 805

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 865

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 925

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 985

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1045

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1105

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1165

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1166 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1225

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              |||||||||||||||||||||||||||||||||||
Db       1226 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 8
I16100
LOCUS       I16100                  1749 bp    DNA     linear   PAT 03-APR-1996
DEFINITION  Sequence 2 from patent US 5474901.
ACCESSION   I16100
VERSION     I16100.1  GI:1251008
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Antibodies to human carboxypeptidase B and methods of use thereof
  JOURNAL   Patent: US 5474901-A 2 12-DEC-1995;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:                4.23e-162      Length:          1749
Score:                    1732.00        Matches:         338
Percent Similarity:       86.01%         Conservative:    0
Best Local Similarity:    86.01%         Mismatches:      0
Query Match:              96.33%         Indels:          55
DB:                       6              Gaps:            2

US-09-980-881A-3 (1-338) x I16100 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286
```

```
Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              |||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 9
I33526
LOCUS       I33526                1749 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 2 from patent US 5593674.
ACCESSION   I33526
VERSION     I33526.1  GI:1824317
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Plasma carboxypeptidase
  JOURNAL   Patent: US 5593674-A 2 14-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:                4.23e-162      Length:           1749
Score:                    1732.00        Matches:          338
Percent Similarity:       86.01%         Conservative:     0
Best Local Similarity:    86.01%         Mismatches:       0
Query Match:              96.33%         Indels:           55
DB:                       6              Gaps:             2

US-09-980-881A-3 (1-338) x I33526 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826
```

```
Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 10
BT006936
LOCUS       BT006936               1272 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA,
            complete cds.
ACCESSION   BT006936
VERSION     BT006936.1  GI:30582710
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00156X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .1272
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma, carboxypeptidase
                     U)"
                     /protein_id="AAP35582.1"
                     /db_xref="GI:30582711"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      373 a    268 c    273 g    358 t
ORIGIN

Alignment Scores:
Pred. No.:              8.73e-162       Length:         1272
Score:                  1727.00         Matches:        337
Percent Similarity:     85.75%          Conservative:   0
Best Local Similarity:  85.75%          Mismatches:     1
Query Match:            96.05%          Indels:         55
DB:                     9               Gaps:           2

US-09-980-881A-3 (1-338) x BT006936 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 666

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
```

```
                                                     ||||||||||||||||||||||||
Db        667 GTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                 ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1207 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 11
BT007990
LOCUS       BT007990                1272 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
            carboxypeptidase U) mRNA, partial cds.
ACCESSION   BT007990
VERSION     BT007990.1  GI:30584818
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone: 'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00156L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .>1272
                     /note="Mutations: 1271:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens carboxypeptidase B2 (plasma,
                     carboxypeptidase U)"
                     /protein_id="AAP36662.1"
                     /db_xref="GI:30584819"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNVL"
BASE COUNT      372 a    268 c    273 g    359 t
ORIGIN

Alignment Scores:
Pred. No.:              8.73e-162       Length:         1272
Score:                  1727.00         Matches:        337
Percent Similarity:     85.75%          Conservative:   0
Best Local Similarity:  85.75%          Mismatches:     1
Query Match:            96.05%          Indels:         55
DB:                     12              Gaps:           2

US-09-980-881A-3 (1-338) x BT007990 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
```

```
                ||||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 666

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  ||||||||||||||||||||||
Db        667 GTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                       ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
                ||||||||||||||||||||||||||||||||||||
Db       1207 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243
```

RESULT 12
BC007057

```
LOCUS       BC007057                1715 bp    mRNA    linear   PRI 12-JUL-2001
DEFINITION  Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495
            IMAGE:3934520, mRNA, complete cds.
ACCESSION   BC007057
VERSION     BC007057.1  GI:13937896
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1715)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 16 Row: f Column: 21
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4503004.
FEATURES             Location/Qualifiers
     source          1. .1715
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="LocusID:1361"
                     /db_xref="taxon:9606"
                     /clone="MGC:12495 IMAGE:3934520"
                     /tissue_type="Skeletal Muscle"
                     /clone_lib="NIH_MGC_81"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
     CDS             25. .1296
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma)"
                     /protein_id="AAH07057.1"
                     /db_xref="GI:13937897"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      518 a    348 c    341 g    508 t
ORIGIN
```

Alignment Scores:

Pred. No.:	1.29e-161	Length:	1715
Score:	1727.00	Matches:	337
Percent Similarity:	85.75%	Conservative:	0
Best Local Similarity:	85.75%	Mismatches:	1
Query Match:	96.05%	Indels:	55
DB:	9	Gaps:	2

US-09-980-881A-3 (1-338) x BC007057 (1-1715)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         91 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 150

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        151 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 210

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        211 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 270

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        271 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 330

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        331 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 390

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        391 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 450

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
```

```
Db        451 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 510

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Db        511 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 570

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175

Db        571 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 630

Qy        175 ------------------------------------------------------------ 175

Db        631 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 690

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183

Db        691 GTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 750

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203

Db        751 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 810

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223

Db        811 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 870

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243

Db        871 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 930

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263

Db        931 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 990

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283

Db        991 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1050

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303

Db       1051 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1110

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323

Db       1111 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1170

Qy        324 ----------------------------------------------------ThrSerAs 326

Db       1171 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1230

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338

Db       1231 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1267
```

```
RESULT 13
AB021968
LOCUS       AB021968               1421 bp    mRNA    linear   ROD 27-JUL-2000
DEFINITION  Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION   AB021968
VERSION     AB021968.1  GI:9558447
KEYWORDS    carboxypeptidase R.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Obata,K., Okada,N.,
            Campbell,W. and Okada,H.
  TITLE     Pro-carboxypeptidase R is an acute phase protein in the mouse,
            whereas carboxypeptidase N is not
  JOURNAL   J. Immunol. 165 (2), 1053-1058 (2000)
  MEDLINE   20341711
   PUBMED   10878383
REFERENCE   2  (bases 1 to 1421)
  AUTHORS   Sato,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-DEC-1998) Tomoo Sato, Fukushimura Hospital, Choju
            Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi
            441-8124, Japan (E-mail:tomoo@tcp-ip.or.jp, Tel:81-532-46-7511,
            Fax:81-532-46-4899)
FEATURES             Location/Qualifiers
     source          1. .1421
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
     CDS             17. .1285
                     /codon_start=1
                     /product="carboxypeptidase R"
                     /protein_id="BAB03402.1"
                     /db_xref="GI:9558448"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
     polyA_site      1421
                     /note="11 a nucleotides"
BASE COUNT      418 a    320 c    305 g    378 t
ORIGIN
```

Alignment Scores:

Pred. No.:	1.39e-138	Length:	1421
Score:	1494.00	Matches:	292
Percent Similarity:	79.64%	Conservative:	21
Best Local Similarity:	74.30%	Mismatches:	25
Query Match:	83.09%	Indels:	55
DB:	10	Gaps:	2

US-09-980-881A-3 (1-338) x AB021968 (1-1421)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Db         80 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 139

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Db        140 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 199

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Db        200 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 259

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Db        260 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 319

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Db        320 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 379

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Db        380 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 439

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Db        440 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 499

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Db        500 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 559
```

```
Qy       161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------- 175
Db       560 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 619

Qy       175 ---------------------------------------------------------- 175

Db       620 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 679

Qy       176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||
Db       680 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAACCGC 739

Qy       184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
Db       740 TCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCCAAA 799

Qy       204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
Db       800 CACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGACTTTAT 859

Qy       224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
Db       860 CCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATCGACCAC 919

Qy       244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
Db       920 ATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTATTCCTAT 979

Qy       264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
Db       980 AACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCAGTTCGT 1039

Qy       284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
Db      1040 GCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAAAGTTTA 1099

Qy       304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
Db      1100 TATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1159

Qy       324 ----------------------------------------------------ThrSerAs 326
                                                                 ||||||||
Db      1160 ACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATACATCAAA 1219

Qy       326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
Db      1220 CCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1256
```

```
RESULT 14
AF164524
LOCUS       AF164524               1490 bp    mRNA    linear   ROD 05-APR-2000
DEFINITION  Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi)
            mRNA, complete cds.
ACCESSION   AF164524
VERSION     AF164524.1  GI:7416966
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1490)
  AUTHORS   Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van
            Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
  TITLE     Characterization of mouse thrombin-activatable fibrinolysis
            inhibitor
  JOURNAL   Thromb. Haemost. 83 (2), 297-303 (2000)
  MEDLINE   20201996
   PUBMED   10739389
REFERENCE   2  (bases 1 to 1490)
  AUTHORS   Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
            Utrecht 3584 CX, The Netherlands
```

```
FEATURES             Location/Qualifiers
     source          1. .1490
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:477035"
     gene            1. .1490
                     /gene="Tafi"
     CDS             84. .1352
                     /gene="Tafi"
                     /function="carboxypeptidase"
                     /codon_start=1
                     /product="thrombin-activatable fibrinolysis inhibitor"
                     /protein_id="AAF62385.1"
                     /db_xref="GI:7416967"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFPVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      438 a    337 c    322 g    393 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.47e-138      Length:          1490
Score:                    1494.00        Matches:         292
Percent Similarity:       79.64%         Conservative:    21
Best Local Similarity:    74.30%         Mismatches:      25
Query Match:              83.09%         Indels:          55
DB:                       10             Gaps:            2

US-09-980-881A-3 (1-338) x AF164524 (1-1490)

Qy         1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db       147 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 206

Qy        21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
Db       207 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 266

Qy        41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
Db       267 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 326

Qy        61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
Db       327 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 386

Qy        81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
Db       387 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 446

Qy       101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
Db       447 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 506

Qy       121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
Db       507 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 566

Qy       141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
Db       567 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 626

Qy       161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------- 175
Db       627 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 686

Qy       175 ---------------------------------------------------------- 175
```

```
Db        687 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 746

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||||
Db        747 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAACCGC 806

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::   ||||||  |||:::||||||||||||||||||||||||||||||||||
Db        807 TCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCCAAA 866

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        867 CACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGACTTTAT 926

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||:::
Db        927 CCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATCGACCAC 986

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||  |||:::||||||||||||||||
Db        987 ATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTATTCCTAT 1046

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1047 AACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCAGTTCGT 1106

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              |||||||||     :::||||||||||||||||||||||||  |||||||||:::|||
Db       1107 GCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAAAGTTTA 1166

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db       1167 TATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1226

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1227 ACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATACATCAAA 1286

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||   |||:::||||||   |||   |||||||||
Db       1287 CCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1323

RESULT 15
AF186188
LOCUS       AF186188               1430 bp    mRNA    linear   ROD 01-OCT-1999
DEFINITION  Mus musculus carboxypeptidase U mRNA, complete cds.
ACCESSION   AF186188
VERSION     AF186188.1  GI:6003651
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Isolation and characterization of mouse liver carboxypeptidase B
            gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Division of Hematology, Washington
            University School of Medicine, 216 S. Kingshighway, St. Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="B6/CBAF1J"
                     /db_xref="taxon:10090"
     CDS             20. .1288
                     /note="pro-carboxypeptidase-B; thrombin-activatable
                     fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
                     plasminogen"
                     /codon_start=1
                     /product="carboxypeptidase U"
                     /protein_id="AAF00528.1"
                     /db_xref="GI:6003652"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVLTEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIGSFNKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      421 a    320 c    310 g    379 t
ORIGIN

Alignment Scores:
Pred. No.:              1.37e-137      Length:          1430
Score:                  1484.00        Matches:         290
Percent Similarity:     79.39%         Conservative:    22
Best Local Similarity:  73.79%         Mismatches:      26
Query Match:            82.54%         Indels:          55
DB:                     10             Gaps:            2

US-09-980-881A-3 (1-338) x AF186188 (1-1430)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||:::||||||||||||||||||||||||||||:::|||
Db         83 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 142

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||:::||||||||||||||||||||||:::   |||
Db        143 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 202

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                 |||||||||:::||||||||||||||||||||||||||||||||:::|||||||||
Db        203 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 262

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||   ||||||   :::||||||:::   :::|||||||||||||||
Db        263 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 322

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::||||||     |||||||||||||||||||||||||||||||||||||||||||||
Db        323 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 382

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||   :::||||||:::|||||||||
Db        383 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCTTAACTGAACAGCATCCTGAC 442

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||   ||||||:::|||||||||||||||||||||||||||||||||||||||||
Db        443 ATGCTCCAGAAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 502

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        503 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 562

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||||||||||||||||||||||||||:::
Db        563 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 622

Qy        175 ------------------------------------------------------------ 175

Db        623 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 682

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||||
Db        683 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAACCGC 742

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::   ||||||  |||:::||||||||||||||||||||||||||||||||||
```

```
Db        743 TCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCCAAA 802

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGACTTTAT 862

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||    ||||||||||||||||:::
Db        863 CCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATCGACCAC 922

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||    |||:::||||||||||||||||
Db        923 ATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTATTCCTAT 982

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 AACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCAGTTCGT 1042

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||         :::||||||||||||||||||||||||   |||||||||:::|||
Db       1043 GCAATTGGAAGTTTTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAAAGTTTA 1102

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db       1103 TATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1162

Qy        324 --------------------------------------------------ThrSerAs 326
                                                               ||||||||
Db       1163 ACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATACATCAAA 1222

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||   |||:::||||||   |||   |||||||||
Db       1223 CCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1259
```

Search completed: January 2, 2004, 00:16:27
Job time : 5247.6 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:19:56 ; Search time 57.1566 Seconds
(without alignments)
1526.015 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1784	99.2	360	4	Q9P2Y6	Q9p2y6 homo sapien
2	1699.5	94.5	423	4	Q15114	Q15114 homo sapien
3	1694.5	94.2	423	4	Q96IY4	Q96iy4 homo sapien
4	1487.5	82.7	422	11	Q9JHH6	Q9jhh6 mus musculu
5	1477.5	82.2	422	11	Q9QZF0	Q9qzf0 mus musculu
6	1466.5	81.6	422	11	Q9EQV9	Q9eqv9 rattus norv
7	916	50.9	198	4	Q9NTI8	Q9nti8 homo sapien
8	697.5	38.8	416	13	Q9PUF2	Q9puf2 bothrops ja
9	688	38.3	408	13	Q8AXN3	Q8axn3 paralichthy
10	646.5	36.0	416	6	Q9XSP3	Q9xsp3 sus scrofa
11	627	34.9	412	11	P97597	P97597 rattus norv
12	610	33.9	437	4	Q8N4T0	Q8n4t0 homo sapien
13	606.5	33.7	434	13	Q8AVS2	Q8avs2 xenopus lae
14	587	32.6	419	13	Q8UUK1	Q8uuk1 gallus gall
15	569	31.6	374	4	Q8IVL8	Q8ivl8 homo sapien
16	558.5	31.1	436	11	Q8R4H4	Q8r4h4 mus musculu
17	528.5	29.4	419	13	Q8AXN4	Q8axn4 paralichthy
18	527.5	29.3	436	4	Q8WXQ8	Q8wxq8 homo sapien
19	521.5	29.0	419	6	Q9TV85	Q9tv85 sus scrofa
20	513.5	28.6	420	11	Q8BMK6	Q8bmk6 mus musculu
21	489	27.2	452	13	Q8UW45	Q8uw45 fugu rubrip
22	484.5	26.9	262	11	Q8BVD0	Q8bvd0 mus musculu
23	477.5	26.6	247	4	Q96KZ9	Q96kz9 homo sapien
24	466.5	25.9	419	13	Q8AXN5	Q8axn5 paralichthy
25	463	25.8	1132	5	Q9W475	Q9w475 drosophila
26	461.5	25.7	422	5	Q9VLZ2	Q9vlz2 drosophila
27	449.5	25.0	226	4	Q8TDE8	Q8tde8 homo sapien
28	444.5	24.7	351	5	Q8T0Q1	Q8t0q1 drosophila
29	443	24.6	584	5	Q9TZH1	Q9tzh1 caenorhabdi
30	441	24.5	351	4	Q8NA08	Q8na08 homo sapien
31	440	24.5	424	5	Q9VL87	Q9vl87 drosophila
32	440	24.5	428	5	Q8IGC4	Q8igc4 drosophila
33	426	23.7	445	5	Q9W478	Q9w478 drosophila
34	419.5	23.3	416	5	Q9VRZ3	Q9vrz3 drosophila
35	416	23.1	488	5	Q9BL88	Q9bl88 caenorhabdi
36	409	22.7	424	5	Q9VS66	Q9vs66 drosophila
37	408.5	22.7	419	5	Q9VLZ1	Q9vlz1 drosophila
38	406.5	22.6	427	5	Q9U9K2	Q9u9k2 aedes aegyp
39	406.5	22.6	430	5	Q9VL86	Q9vl86 drosophila
40	403	22.4	323	5	Q19121	Q19121 caenorhabdi
41	401	22.3	540	5	Q9TZH2	Q9tzh2 caenorhabdi
42	399.5	22.2	430	5	Q8SZP6	Q8szp6 drosophila
43	398	22.1	453	5	Q9V342	Q9v342 drosophila
44	392	21.8	423	5	O61532	O61532 drosophila
45	391.5	21.8	454	5	Q9TZC6	Q9tzc6 caenorhabdi

ALIGNMENTS

RESULT 1
Q9P2Y6
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; AB011969; BAA90475.1; -.
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

this art has 37aa insert see #2

Query Match 99.2%; Score 1784; DB 4; Length 360;
Best Local Similarity 99.1%; Pred. No. 6.2e-154;
Matches 335; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQTGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 180
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWR 202

Qy        181 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 262

Qy        241 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 300
              |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        263 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGS 322

Qy        301 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 338
              ||||||||||||||||||||||||||||||||||||||
Db        323 ETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKLLPLSLK 360
```

RESULT 2
Q15114
ID Q15114 PRELIMINARY; PRT; 423 AA.
AC Q15114;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE PCPB protein.
GN PCPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; M75106; AAA60042.1; -.
DR HSSP; P00730; 1PYT.
DR MEROPS; M14.009; -.
DR Genew; HGNC:2300; CPB2.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
FT CHAIN 23 423 PLASMA CARBOXYPEPTIDASE B.
SQ SEQUENCE 423 AA; 48442 MW; 9B383272F6EE79F4 CRC64;

Query Match 94.5%; Score 1699.5; DB 4; Length 423;
Best Local Similarity 89.8%; Pred. No. 3.8e-146;
Matches 324; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                           ||||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383
```

RESULT 3
Q96IY4
ID Q96IY4 PRELIMINARY; PRT; 423 AA.
AC Q96IY4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B2 (Plasma).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC007057; AAH07057.1; -.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 423 AA; 48412 MW; 9B383E03F6EE7CF5 CRC64;

Query Match 94.2%; Score 1694.5; DB 4; Length 423;
Best Local Similarity 89.5%; Pred. No. 1.1e-145;
Matches 323; Conservative 0; Mismatches 1; Indels 37; Gaps 1;

```
Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFY 202

Qy        176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                           ||||||||||||||||||||||||||||||
Db        203 GIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASK 262

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 322

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        323 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 382

Qy        324 T 324
              |
Db        383 T 383


RESULT 4
Q9JHH6
ID   Q9JHH6      PRELIMINARY;      PRT;   422 AA.
AC   Q9JHH6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE   (1110032P04Rik protein).
GN   CPB2 OR TAFI OR 1110032P04RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20341711; PubMed=10878383;
RA   Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA   Campbell W., Okada H.;
RT   "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT   whereas carboxypeptidase N is not.";
RL   J. Immunol. 165:1053-1058(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20201996; PubMed=10739389;
RA   Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA   van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT   "Characterization of mouse thrombin-activatable fibrinolysis
RT   inhibitor.";
RL   Thromb. Haemost. 83:297-303(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AB021968; BAB03402.1; -.
DR   EMBL; AF164524; AAF62385.1; -.
DR   EMBL; AK004045; BAB23141.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48870 MW;  99113755669D55CB CRC64;

  Query Match           82.7%;  Score 1487.5;  DB 11;  Length 422;
  Best Local Similarity   78.1%;  Pred. No. 7.4e-127;
  Matches  282;  Conservative   20;  Mismatches   22;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||:||||||||||:|||||||||:||||||||: | |||:||||||||||:|||
Db         22 FQSGQVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKA 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||| || :||: :|||||:||  |||||||||||||||||||||||||||||:|||
Db         82 HLNVSRIPFNVLMNNVEDLIEQQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPD 141

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              || ||:|||||||||||||||  |||||||||||||||||||||||||:
Db        142 MLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFH 201

Qy        176 ------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                           |||||||| : || |:||||||||||
Db        202 GKENLYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASK 261

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||:|||||||||||||||||||||| ||||||: |||||||||||| |:|||||
Db        262 HWCEKGASSSSCSETYCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSY 321

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              |||||||||||||||||||||  :|||||||| |||:|||||| ||||||||||||
Db        322 NRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSF 381

Qy        324 T 324
              |
Db        382 T 382


RESULT 5
Q9QZF0
ID   Q9QZF0      PRELIMINARY;      PRT;   422 AA.
AC   Q9QZF0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase U.
GN   CPB2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/CBAF1J;
RA   He Y.C., Broze G.;
RT   "Isolation and characterization of mouse liver carboxypeptidase B
RT   gene.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF186188; AAF00528.1; -.
DR   HSSP; P00730; 5CPA.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48832 MW;  B53FFFB09943954E CRC64;

  Query Match           82.2%;  Score 1477.5;  DB 11;  Length 422;
  Best Local Similarity   77.6%;  Pred. No. 6e-126;
  Matches  280;  Conservative   21;  Mismatches   23;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||||||:||||||||||:|||||||||:||||||||: | |||:||||||||||:|||
```

```
Db         22 FQSGQVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKA 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||| || :||: :|||||:||  |||||||||||||||||||||||||| :||:|||
Db         82 HLNVSRIPFNVLMNNVEDLIEQQTFNDTVSPRASASYYEQYHSLNEIYSWIEVLTEQHPD 141

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              || ||:|||||||||||||||||  |||||||||||||||||||||||||:
Db        142 MLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFH 201

Qy        176 --------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                              ||||||||| : || |:||||||||||
Db        202 GKENLYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASK 261

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||:|||||||||||||||||||||||| ||||||: |||||||||||| |:|||||
Db        262 HWCEKGASSSSCSETYCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSY 321

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||   :|||||||| |||:||||||| ||||||||||||||
Db        322 NRSKSKDHEELSLVASEAVRAIGSFNKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSF 381

Qy        324 T 324
              |
Db        382 T 382


RESULT 6
Q9EQV9
ID   Q9EQV9      PRELIMINARY;      PRT;   422 AA.
AC   Q9EQV9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Pre-procarboxypeptidase R.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20471387; PubMed=11021404;
RA   Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA   Okada N., Okada H.;
RT   "Molecular cloning and partial characterization of rat
RT   procarboxypeptidase R and carboxypeptidase N.";
RL   Microbiol. Immunol. 44:719-728(2000).
DR   EMBL; AB042598; BAB18617.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   422 AA;  48826 MW;  FFFD32A51A9366C8 CRC64;

  Query Match             81.6%;  Score 1466.5;  DB 11;  Length 422;
  Best Local Similarity   76.7%;  Pred. No. 6.1e-125;
  Matches  277;  Conservative   22;  Mismatches   25;  Indels   37;  Gaps    1;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |||| ||:|||||||||:|||||||:|||||||: | |||:||||||||::|||
Db         22 FQSGHVLSALPRTSRQVQLLQNLTTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVNSVKA 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              :|| | || :||: :|||||||| |||||||||:|||||||||||||||| |||:|||
Db         82 YLNASRIPFNVLMNNVEDLIQQQTSNDTVSPRASSSYYEQYHSLNEIYSWIEVITEQHPD 141

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH----- 175
              || ||:||||:|||||||||||||   |||||||||||||||||||||||:
Db        142 MLQKIYIGSSYEKYPLYVLKVSGKEHRVKNAIWIDCGIHAREWISPAFCLWFIGYVTQFH 201

Qy        176 --------------------------------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                              |||||||| : || |:||||||||||
Db        202 GKENTYTRLLRHVDFYIMPVMNVDGYDYTWKKNRMWRKNRSVHMNNRCVGTDLNRNFASK 261

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              ||||:|||| |||||||||||||||||| |||||| |||||||||||| |:|||||
Db        262 HWCEKGASSFSCSETYCGLYPESEPEVKAVADFLRRNINHIKAYISMHSYSQQILFPYSY 321

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              ||||||||||||||||||||  :|||||||| |||:||||||| ||||||||||||||
Db        322 NRSKSKDHEELSLVASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSF 381

Qy        324 T 324
              |
Db        382 T 382


RESULT 7
Q9NTI8
ID   Q9NTI8      PRELIMINARY;      PRT;   198 AA.
AC   Q9NTI8;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   BA139H14.2 (Carboxypeptidase B2 (Plasma)) (Fragment).
GN   CPB2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tromans A.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL137141; CAB92622.1; -.
DR   HSSP; P00730; 1CPX.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
KW   Carboxypeptidase.
FT   NON_TER     198    198
SQ   SEQUENCE   198 AA;  22460 MW;  32F005305621C2A5 CRC64;

  Query Match             50.9%;  Score 916;  DB 4;  Length 198;
  Best Local Similarity   99.4%;  Pred. No. 2.7e-75;
  Matches  174;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         23 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 82

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 142

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGH 175
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db        143 MLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGH 197


RESULT 8
Q9PUF2
ID   Q9PUF2      PRELIMINARY;      PRT;   416 AA.
AC   Q9PUF2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase homolog.
OS   Bothrops jararaca (Jararaca).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
OX   NCBI_TaxID=8724;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT   "Screening of Bothrops jararaca pancreas cDNA library.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF190274; AAF01344.1; -.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ   SEQUENCE   416 AA;  47723 MW;  4F999854DD72B7A7 CRC64;

  Query Match             38.8%;  Score 697.5;  DB 13;  Length 416;
  Best Local Similarity   38.8%;  Pred. No. 5.7e-55;
  Matches  140;  Conservative   66;  Mismatches  112;  Indels   43;  Gaps    5;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|    ||   :|  | |    ::  |:| : :|:  :  | | : |     |::
Db         22 FDGEKVYRVTPRNEDEVYFLNYLANIVQVDFWRPDSVELVKAEMTVDFRIEADRCSEVES 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  ||:   :|: ::: :: :|: |     | :   ||:|:|  :| :|   |   :|
Db         82 ILQQSGLNYEILIDNLQAVLDRQLDNHA---RTAGYNYEKYNSWEKIDAWTADIANENPS 138

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              ::::: ||::||  |: :||| ||    | ||:|||| |||||||||| ||:
Db        139 LVSRLQIGTTFEGRPMPLLKV-GKPGVNKKAIFIDCGFHAREWISPAFCQWFVREAVRTY 197

Qy        174 ----------------------------GHNRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                         :|||| ||  | : |||| |||| :
Db        198 GKETIMTQLLNKLDFYILPVLNIDGYVYSWKQSRMWRKTRSVNAGSTCIGTDPNRNFDAA 257

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||  ||| : |||||| |||| | ||:| |:||| : |:||:::|||| :::|||
Db        258 -WCSVGASRNPCSETYCGSKPESEKETKALADFIRRNRSIIQAYLTIHSYSQMLLYPYSY 316

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              |   : ::::|: :| ||:| : |    | ||:| |: |:| | || ||| || ||||:|
Db        317 TYDLTSNNKKLNSIAKEAIREL-KVLFGTEYTYGPGAATIYPAAGGSDDWAYDQGIKYAF 375

Qy        324 T 324
              |
Db        376 T 376


RESULT 9
Q8AXN3
ID   Q8AXN3      PRELIMINARY;      PRT;   408 AA.
AC   Q8AXN3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B (Fragment).
GN   CPB.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Srivastava A.S., Kurokawa T., Suzuki T.;
RT   "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT   A2 and B from the Japanese flounder, Paralichthys olivaceus.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB099302; BAC53789.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   408 AA;  45904 MW;  809CAACD9D2DA34B CRC64;

  Query Match             38.3%;  Score 688;  DB 13;  Length 408;
  Best Local Similarity   39.6%;  Pred. No. 4e-54;
  Matches  143;  Conservative   61;  Mismatches  115;  Indels   42;  Gaps    5;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|    |      | ::::|  : |:  |:| :: |:     |   | |  :| |
Db         12 FEGEKVFRLKPVFDEHVTLIRDLANSIEVDFWRPESSQLVTIDIDVDIRVPAIYLDMVYT 71

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  | :   ||: |::  :  :     : |   |  | :|::|::: |||  |:  :||
Db         72 LLQQSDMEHEVLIEDLQSAVDAEAD---LKPSPRAHSYTKYNNLDKVQSWIASISSSNPD 128

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              :::|  ||:::|  |: |||:  |  : | ||::||||||||||||||| ||:
Db        129 LISKQVIGNTYEGRPMTVLKLGKKSSSTKPAIFMDCGIHAREWISPAFCQWFVKEALSTY 188

Qy        174 ----------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      |||||| ||  : : ||| | ||||
Db        189 GSDSQMTSLLNQMDVFVLPVFNIDGYDFTHKSNRMWRKTRSRKSGSSCIGADPNRNF-DA 247

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||  ||| : ||:|:||  |||| ||| || |:||| : ||||:::||||| ::|||||
Db        248 GWCSMGASDNPCSDTFCGYTPESEIEVKNVADFIRRNRSIIKAYLTVHSYSQLLLFPYSY 307

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
              |   : || ||  ||  |  |: ::   | || | |: |:| | || ||| ||||:|||:
Db        308 TYQLAADHSELMKVAEGASAAL-RSLYGTTYTSGPGATTIYPAAGGSDDWAYDLGVKYSY 366

Qy        324 T 324
              |
Db        367 T 367


RESULT 10
Q9XSP3
ID   Q9XSP3      PRELIMINARY;      PRT;   416 AA.
AC   Q9XSP3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Procarboxypeptidase B precursor (EC 3.4.17.2).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=99321926; PubMed=10391940;
RA   Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA   Hershberger C., Aviles F.;
RT   "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT   Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL   J. Biol. Chem. 274:19925-19933(1999).
DR   EMBL; AJ133775; CAB46991.1; -.
DR   HSSP; P09955; 1PBA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Hydrolase; Signal.
FT   SIGNAL        1     15       POTENTIAL.
```

```
FT   CHAIN        16    416       PROCARBOXYPEPTIDASE B.
SQ   SEQUENCE   416 AA;  47381 MW;  5FA7FCED7B45AE6A CRC64;

  Query Match           36.0%;  Score 646.5;  DB 6;  Length 416;
  Best Local Similarity   38.0%;  Pred. No. 2.5e-50;
  Matches  137;  Conservative   60;  Mismatches  121;  Indels   43;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |:  :|        : :|  | :| :|  |:| :   |     | | | | |:  |:
Db         22 FEGEKVFRVNVEDENDISLLHELASTRQIDFWKPDSVTQIKPHSTVDFRVKAEDILAVED 81

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |  : :   ||: ::  ::: |  :     | :   ||:|::   | :| : :|  :||
Db         82 FLEQNELQYEVLINNLRSVLEAQFDSRV---RTTGHSYEKYNNWETIEAWTKQVTSENPD 138

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
               ::::  ||::|    :|:||| ||     | ||::||| ||||||| ||| ||:
Db        139 LISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCGFHAREWISHAFCQWFVREAVLTY 197

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                       |:      |||||| ||   |   ||||| ||||
Db        198 GYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWRKTRSTNAGTTCIGTDPNRNF-DA 256

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               ||  |||:  | |||||   ||| | ||:| |:| |:: ||||:::|||||  |::||||
Db        257 GWCTTGASTDPCDETYCGSAAESEKETKALADFIRNNLSSIKAYLTIHSYSQMILYPYSY 316

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                  ::: ||: :|  ||: :  |   |:||:| |: |:| | || ||| || ||||||
Db        317 DYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGATTIYPAAGGSDDWAYDQGIKYSF 375

Qy        324 T 324
               |
Db        376 T 376


RESULT 11
P97597
ID   P97597      PRELIMINARY;      PRT;   412 AA.
AC   P97597;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (Fragment).
GN   R-CPA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague Dawley;
RX   MEDLINE=97149430; PubMed=8996238;
RA   Lutzelschwab C., Pejler G., Aveskogh M., Hellman L.;
RT   "Secretory granule proteases in rat mast cells. Cloning of 10
RT   different serine proteases and a carboxypeptidase A from various rat
RT   mast cell populations.";
RL   J. Exp. Med. 185:13-29(1997).
DR   EMBL; U67914; AAB48267.1; -.
DR   HSSP; P09955; 1NSA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Signal.
FT   NON_TER       1      1
FT   SIGNAL       <1     10       POTENTIAL.
FT   CHAIN        11    412       MAST CELL CARBOXYPEPTIDASE A.
SQ   SEQUENCE   412 AA;  47944 MW;  DAB59555FC49137D CRC64;

  Query Match           34.9%;  Score 627;  DB 11;  Length 412;
  Best Local Similarity   37.5%;  Pred. No. 1.5e-48;
  Matches  135;  Conservative   55;  Mismatches  128;  Indels   42;  Gaps    6;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKA 60
              |   :|      : :|  :|:||| | |:  | |      |      | | |   :   :::
Db         16 FDREKVFRVKLQDEKQASILKNLTQTIELDFWYPDAIHDIAVNMTVDFRVTEKESQTIQS 75

Qy         61 HLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPD 120
               |    :   :|: |::: | :|   |      |     | :|:  |:| || | : |:||:
Db         76 TLEQHKMDYEILINDLQEEIDKQF--DVKEEIAGRHSYAKYNDWNKIVSWTEKMVEKHPE 133

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
               |:::| |||: |  ||||||: |::   : ||::|||||||||:||||| ||:
Db        134 MVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCGIHAREWVSPAFCQWFVYQAAKSY 192

Qy        174 GHN----------------------------RMWRKNRSFYANNHCIGTDLNRNFASK 203
               | |                           ||||||||   |: ||||||||||
Db        193 GKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWRKNRSKNPNSTCIGTDLNRNF-DV 251

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
               |     : : |   | |  |||| | ||| :|:|  ::| |||||: ||||| ::||| |
Db        252 SWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSHLNSIKAYITFHSYSQMLLFPYGY 311

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
               |     :|::|  ||  |   :  :   ||| :|   : |:|    |   || ||||||::|
Db        312 TIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIASTIYKTSGSSLDWAYDLGIKHTF 370


RESULT 12
Q8N4T0
ID   Q8N4T0      PRELIMINARY;      PRT;   437 AA.
AC   Q8N4T0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033684; AAH33684.1; -.
DR   Genew; HGNC:17245; CPA6.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   437 AA;  51024 MW;  9970680D429A0A98 CRC64;

  Query Match           33.9%;  Score 610;  DB 4;  Length 437;
  Best Local Similarity   35.4%;  Pred. No. 5.6e-47;
  Matches  129;  Conservative   62;  Mismatches  127;  Indels   46;  Gaps    7;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVK--KKQVHFFVNASDVDNV 58
              :   :|:  :|:|  :   |: ::   :: |||| :   : :     ||   | |    :
Db         37 YAGDKVIRFIPKTEEEAYALKKISYQLKVDLWQPSSISYVSEGTVTDVHIPQNGSRA--L 94

Qy         59 KAHLNVSGIPCSVLLADVEDLIQQQISNDTVSPRASAS--YYEQYHSLNEIYSWIEFITE 116
               | |  : |   ||: |::  :::  |  |   | | |   || |||| || :|:  : :
Db         95 LAFLQEANIQYKVLIEDLQKTLEKGSSLHTQRNRRSLSGYNYEVYHSLEEIQNWMHHLNK 154

Qy        117 RHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI--- 173
               |  ::    || |:|   |::||: |:   | |:||||||||||| |||| ||:
Db        155 THSGLIHMFSIGRSYEGRCLFILKL-GRRSRLKRAVWIDCGIHAREWIGPAFCQWFVKEA 213
```

```
Qy        174 ---------------------------------GHNRMWRKNRSFYANNHCIGTDLNRN 199
                                                ::| ||| ||  :   | | | |||
Db        214 LLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSWTNDRFWRKTRSRNSRFRCRGVDANRN 273

Qy        200 FASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVF 259
              :  | ||:||||   | :|||| :||||||||||:|||::   |:||:| |:|:| :::
Db        274 WKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAVANFLRKHRKHIRAYLSFHAYAQMLLY 332

Qy        260 PYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGI 319
              ||||  :   :   :   | :|| |::      || :|  | |||:: |   || |  ||
Db        333 PYSYKYATIPNFRCVESAAYKAVNALQSV-YGVRYRYGPASTTLYVSSGSSMDWAYKNGI 391

Qy        320 KYSF 323
               |:|
Db        392 PYAF 395


RESULT 13
Q8AVS2
ID   Q8AVS2      PRELIMINARY;      PRT;   434 AA.
AC   Q8AVS2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041315; AAH41315.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   434 AA;  50581 MW;  B42BD77843910591 CRC64;

  Query Match             33.7%;  Score 606.5;  DB 13;  Length 434;
  Best Local Similarity   36.4%;  Pred. No. 1.1e-46;
  Matches  131;  Conservative   60;  Mismatches  124;  Indels   45;  Gaps    8;

Qy          5 QVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKK--KQVHFFVNASDVDNVKAHL 62
              :|:   |:  |:   |:::    :: |||| :   | |      ||   |:|:|  : :|
Db         37 KVIRVFPPKNEREANELKDMYQQLQVDLWQPSSISHIGKDTVTDVHTSGNSSEV--LLTYL 94

Qy         63 NVSGIPCSVLLADVEDLIQ-QQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPD 120
                : |   :|: :|: ::: ||        |: : | | :|| |:|| ||: :: : | |
Db         95 TKAKIQHKILVNNVQSMLEAQQAFRPRRKRRSLSRYNYNEYHPLHEIESWMFYMNKTHHD 154

Qy        121 MLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------- 173
              :::   || |:|   |||||:    :: | |||||||:|||||| |||| ||:
Db        155 LVSLFTIGKSYEGRSLYVLKLGKDTKSYKKAIWIDCGMHAREWIGPAFCQWFVKEAINSY 214

Qy        174 ------------------------GH------NRMWRKNRSFYANNHCIGTDLNRNFASK 203
                                     |:      :| ||| ||      | | | |||:
Db        215 NTDPAMKKILNLLYIYVMPVFNVDGYHYSWHSDRFWRKTRSKNTRYQCYGVDANRNW-KV 273

Qy        204 HWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSY 263
              || :|||| : |  |||| | |||||||||| || :    ::||:| |:|:| :::||||
Db        274 HWSDEGASLNPCDNTYCGPYAESEPEVKAVAQFLYKQRKHVRAYMSFHAYAQMLLYPYSY 333

Qy        264 TRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSF 323
                    :   :   | || || :::   || ||  | ||||  |    || |: || ||:
Db        334 QYGAIPNFGCVESAAHNAVLAI-RSAYGIRYRHGPASSTLYLTSGSSMDWAYNNGIPYSY 392


RESULT 14
Q8UUK1
ID   Q8UUK1      PRELIMINARY;      PRT;   419 AA.
AC   Q8UUK1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Preprocarboxypeptidase A precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=domesticus; TISSUE=Pancreas;
RA   Hasegawa S., Honda K., Hikami Y.;
RT   "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
RT   precursor.";
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X64539; CAA45837.1; -.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Signal; Carboxypeptidase.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN      113    419       CARBOXYPEPTIDASE A.
SQ   SEQUENCE   419 AA;  47006 MW;  299C334274BCBC9C CRC64;

  Query Match             32.6%;  Score 587;  DB 13;  Length 419;
  Best Local Similarity   35.5%;  Pred. No. 6.5e-45;
  Matches  134;  Conservative   65;  Mismatches  106;  Indels   72;  Gaps    9;

Qy          1 FQSGQVLAALPRTSRQVQVLQNLTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDN--- 57
              |   |||  :| :  ::| :| |                  ::  |: |:::   : |
Db         21 FVGHQVLRIVPSSDAELQKVQELQE---------------LEHLQLDFWLSPRGLGNPVD 65

Qy         58 ----------VKAHLNVSGIPCSVLLADVEDLI----QQQISNDTVSPRASASY-YEQYH 102
                        |||||  :|:| |::: ||: |:     | :      | ::::: |  ||
Db         66 IRVPFPSLQPVKAHLEANGVPYSIMIEDVQALVDLEQMQMLRRRRFVPLSTSTFDYTSYH 125

Qy        103 SLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHARE 162
              :| |||:::: :   :|::::|: || : |  |:|||| | |  | : ||||| |||:||
Db        126 TLEEIYAFMDLLVAENPNLVSKLEIGRTTENRPIYVLKFS-KGGTNRPAIWIDTGIHSRE 184

Qy        163 WISPAFCLWFI---------------------------------GHNRMWRKNRSFYA 187
              |:: |  :||                                 ||||| || .:
Db        185 WVTQASGIWFAKKIVQEQDEGLANILDQMDIFLEIVTNPDGFAFTHTSNRMWRKTRSKRS 244

Qy        188 NNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAY 247
               : |:| | |||:    :   ||||: ||||| | |  |||||||:  |:: : |'|||:
Db        245 GSLCVGVDPNRNW-DAGFGGSGASSNPCSETYHGPYANSEPEVKAIVDFVKSHKN-IKAF 302

Qy        248 ISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAP 307
              ||:||||| :::|| || :   | |||  || ||| |:  :   | | :|    |:| |
Db        303 ISIHSYSQLLLYPYGYTTTAVPDKEELHQVAKEAVAAL-SSLYGTNYKYGSIITTIYQAS 361

Qy        308 GGGDDWIYDLGIKYSFT 324
              ||  || |: |||||||
Db        362 GGTIDWTYNQGIKYSFT 378


RESULT 15
Q8IVL8
ID   Q8IVL8      PRELIMINARY;      PRT;   374 AA.
AC   Q8IVL8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zn-carboxypeptidase.
OS   Homo sapiens (Human).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Obaya A.J., Lopez-Otin C.;
RT   "A new Zn-carboxypeptidase highly expressed in ovary.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ422118; CAD19478.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   374 AA;  42529 MW;  404C373BB841AAD2 CRC64;

  Query Match             31.6%;  Score 569;  DB 4;  Length 374;
  Best Local Similarity   39.0%;  Pred. No. 2.4e-43;
  Matches  110;  Conservative   51;  Mismatches   81;  Indels   40;  Gaps    5;

Qy         81 QQQISNDTVSPRASASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVL 139
              :|:| : :||| :  :| |  || : ||| |:  |:|:: :::|:  :| ::| :|:| |
Db         29 RQEIVDKSVSPWSLETYSYNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYL 88

Qy        140 KVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFI------------------------- 173
              |:|       |  ||:||||||||||:|||| ||:
Db         89 KISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLP 148

Qy        174 -----GH------NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGL 222
                  |:      :|:|||:|| : |  | |||||||  :  ||  ||| :   :|:||
Db        149 VLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGT 207

Qy        223 YPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAV 282
               | |||| ||||||:    : |  :::|||| | |: || ||::|| :| |:  |  :|
Db        208 GPVSEPETKAVASFIESKKDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAA 267

Qy        283 RAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 324
               |: |    | |  |  :: || : |   ||  |:|| :|:|
Db        268 NAL-KAKYGTNYRVGSSADILYASSGSSRDWARDIGIPFSYT 308


Search completed: January  1, 2004, 19:31:56
Job time : 58.1566 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:33:22 ; Search time 369.048 Seconds
(without alignments)
2472.334 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135914_26819/app_query.fasta_1.1628
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_577_@runat_31122003_135914_26819 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1785	99.3	1573	21	AAC81962	Human brain carbox
2	1749.5	97.3	1625	25	ABX13670	Human protease cDN
3	1732	96.3	1272	20	AAV74302	Human plasma carbo
4	1732	96.3	1728	24	ABN95899	Gene #2397 used to
5	1732	96.3	1749	17	AAT11671	Human plasma carbo
6	1732	96.3	1749	18	AAT62846	Human plasma carbo
7	1727	96.1	1749	14	AAQ41001	Human plasma carbo
8	902	50.2	1400	21	AAF18005	Lung cancer associ
9	646.5	36.0	1215	16	AAQ90600	Porcine Tyr-His-Me
10	637.5	35.5	1263	17	AAT42494	Human pancreatic c
11	637.5	35.5	1263	19	AAV41795	Human pancreatic c
12	637.5	35.5	1284	17	AAT42506	ProHCPB gene with
13	622	34.6	2023	25	ABZ75116	Anti-human seminal
14	619.5	34.5	1302	25	ABV77254	Nucleotide sequenc
15	619.5	34.5	1311	22	AAD03837	Human carboxypepti
16	619.5	34.5	1993	24	AAD30579	Human protease, PR
17	619.5	34.5	2154	18	AAV17331	PreproHCPB-linker-
18	618	34.4	1332	24	ABL65831	Lung cancer relate
19	610	33.9	927	17	AAT35760	Rat mature carboxy
20	591	32.9	416	25	ABX49430	Bovine EST associa
21	582	32.4	921	16	AAQ90601	Porcine carboxypep
22	571.5	31.8	2128	22	AAD03839	Human carboxypepti
23	569.5	31.7	1378	25	ABT33355	NOVX DNA sequence
24	569	31.6	1125	22	AAH76477	Human secreted met
25	569	31.6	1125	22	AAH76479	cDNA encoding huma
26	569	31.6	1125	24	ABK31745	DNA encoding novel
27	569	31.6	1222	25	ABT33353	NOVX DNA sequence
28	569	31.6	1342	22	AAH76476	Human secreted met
29	569	31.6	1342	22	AAH76478	Human secreted met
30	569	31.6	1344	25	ABQ77393	Human zinc-binding
31	569	31.6	1603	24	AAD39053	Human 23566 (carbo
32	569	31.6	1826	22	AAF81728	Human protease and
33	567	31.5	999	17	AAT42500	Mature HCPB coding
34	567	31.5	1053	17	AAT42497	mature HCPB-(His)6
35	560	31.1	1053	20	AAX24804	Human carboxypepta
36	560	31.1	1059	17	AAT42511	Modified HCPB (D25
37	560	31.1	1059	18	AAT62787	Carboxypeptidase B
38	559	31.1	1059	17	AAT42512	Modified HCPB (D25
39	559	31.1	1059	18	AAT62788	Carboxypeptidase B
40	556	30.9	1059	18	AAT62806	Carboxypeptidase B
41	554	30.8	1059	18	AAT62790	Carboxypeptidase B
42	554	30.8	1059	18	AAT62801	Carboxypeptidase B
43	554	30.8	1200	22	AAD05477	Novel human protea
44	553	30.8	1059	18	AAT62789	Carboxypeptidase B
45	553	30.8	1196	25	ABT33352	NOVX DNA sequence

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.
XX
AC AAC81962;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B cDNA.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             18..1100
FT                   /*tag= a
FT                   /product= "carboxypeptidase B"
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
DR   P-PSDB; AAB11457.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Claim 2b; Page 64-68; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

Alignment Scores:
Pred. No.:              1.31e-186      Length:          1573
Score:                  1785.00        Matches:         335
Percent Similarity:     99.41%         Conservative:    1
Best Local Similarity:  99.11%         Mismatches:      2
Query Match:            99.28%         Indels:          0
DB:                     21             Gaps:            0

US-09-980-881A-3 (1-338) x AAC81962 (1-1573)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         84 TTCCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 143

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        144 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 203

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        204 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 263

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        264 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 323

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        324 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 383

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        384 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 443

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        444 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 503

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        504 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 563

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        564 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 623

Qy        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        624 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGCAACTTT 683

Qy        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        684 GTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 743

Qy        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        744 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 803

Qy        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        804 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 863

Qy        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        864 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 923

Qy        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        924 GCAGTTCGTGCTATTGACAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 983

Qy        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        984 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1043

Qy        321 TyrSerPheThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1044 TATTCGTTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1097

RESULT 2
ABX13670
ID   ABX13670 standard; cDNA; 1625 BP.
XX
AC   ABX13670;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease cDNA.
XX
KW   Human; gene; ss; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..16
FT                   /*tag= a
FT   CDS             17..1177
FT                   /*tag= b
FT                   /product= "Protease"
FT   variation       replace(521,A)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR           1178..1625
FT                   /*tag= d
```

XX
PN US2002137179-A1.
XX
PD 26-SEP-2002.
XX
PF 21-MAR-2001; 2001US-0813133.
XX
PR 21-MAR-2001; 2001US-0813133.
XX
PA (GANW/) GAN W.
PA (KETC/) KETCHUM K A.
PA (DFRA/) DI FRANCESCO V.
PA (BEAS/) BEASLEY E M.
XX
PI Gan W, Ketchum KA, Di Francesco V, Beasley EM;
XX
DR WPI; 2003-102387/09.
DR P-PSDB; ABG72093, ABG72095.
XX
PT New isolated human protease proteins, useful for developing therapeutic
PT or diagnostic compositions, particularly for developing human
PT therapeutic agents that modulate protease activity in cells or tissues
PT -
XX
PS Claim 4; Fig 1; 80pp; English.
XX
CC The invention discloses an isolated human protease peptide, its allelic
CC variant or orthologue. The proper functioning of the cell requires
CC careful control of the levels of important structural proteins, enzymes
CC and regulatory proteins. One of the ways the cell reduces the steady
CC state level of a particular protein is by proteolytic degradation.
CC Proteolysis can also be used to convert a pre or pro-protein in to an
CC active form. Proteases also regulate many different cell proliferation,
CC differentiation and signalling processes. The peptides and nucleic acid
CC molecules are useful in the development of human therapeutics (gene and
CC protein therapy) and diagnostic compositions. The peptides are also
CC useful for raising antibodies or eliciting an immune response (vaccine),
CC as a reagent (including the labeled reagent) in assays designed to
CC quantitatively determine levels of the protein (or its binding partner or
CC ligand) in biological fluids, or as markers for tissues in which the
CC corresponding protein is preferentially expressed and in methods for
CC identifying a modulator of the peptide or an agent that binds to the
CC peptide. The agents identified are useful for treating protease-related
CC conditions that are specific for the subfamily of proteases that the
CC peptide belongs to, particularly in cells and tissues that express the
CC protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC disorders. The modulator of the peptide is also useful for treating a
CC disorder characterised by an absence of, inappropriate or unwanted
CC expression of the protein. The sequence presented is the human protease
CC cDNA, the gene for which is located on chromosome 13.
XX
SQ Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

Alignment Scores:

Pred. No.:	1.14e-182	Length:	1625
Score:	1749.50	Matches:	336
Percent Similarity:	94.38%	Conservative:	0
Best Local Similarity:	94.38%	Mismatches:	2
Query Match:	97.30%	Indels:	18
DB:	25	Gaps:	1

US-09-980-881A-3 (1-338) x ABX13670 (1-1625)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 142

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 202

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 262

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 322

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 382

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 442

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 502

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||
Db        503 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 562

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 622

Qy        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTT 682

Qy        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 GCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 742

Qy        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 802

Qy        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 862

Qy        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 922

Qy        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        923 GCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 982

Qy        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1042

Qy        321 TyrSerPhe---------------------------------------------------- 323
              |||||||
Db       1043 TATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGT 1102

Qy        324 -ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
               ||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 TACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1148
```

RESULT 3
AAV74302
ID AAV74302 standard; cDNA; 1272 BP.
XX
AC AAV74302;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;

```
KW   polymorphism detection; thrombotic disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   P-PSDB; AAW92270.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Example 1; Page 24; 35pp; English.
XX
CC   This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Alignment Scores:
Pred. No.:              6.72e-181       Length:         1272
Score:                  1732.00         Matches:        338
Percent Similarity:     86.01%          Conservative:   0
Best Local Similarity:  86.01%          Mismatches:     0
Query Match:            96.33%          Indels:         55
DB:                     20              Gaps:           2

US-09-980-881A-3 (1-338) x AAV74302 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 666

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  ||||||||||||||||||||
Db        667 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||
Db       1207 CCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAA 1243

RESULT 4
ABN95899
ID   ABN95899 standard; DNA; 1728 BP.
XX
AC   ABN95899;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2397 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
```

OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX
DR WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 2397; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;

Alignment Scores:

Pred. No.:	1.06e-180	Length:	1728
Score:	1732.00	Matches:	338
Percent Similarity:	86.01%	Conservative:	0
Best Local Similarity:	86.01%	Mismatches:	0
Query Match:	96.33%	Indels:	55
DB:	24	Gaps:	2

US-09-980-881A-3 (1-338) x ABN95899 (1-1728)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 505

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 565

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        566 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 625

Qy        175 ------------------------------------------------------------ 175

Db        626 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||
Db        686 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 745

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 805

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 865

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 925

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 985

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1045

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1105

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1165

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1166 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1225

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1226 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262
```

RESULT 5
AAT11671
ID AAT11671 standard; DNA; 1749 BP.
XX
AC AAT11671;
XX
DT 25-MAR-2003 (updated)
DT 12-APR-1996 (first entry)
XX
DE Human plasma carboxypeptidase B coding sequence.
XX
KW Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW purification; plasminogen; affinity column; ss.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag= a
FT                   /product= Human plasma carboxypeptidase B.
FT   sig_peptide     41..106
FT                   /*tag= b
FT   mat_peptide     107..1309
FT                   /*tag= c
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   P-PSDB; AAR90293.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:
Pred. No.:              1.08e-180      Length:         1749
Score:                  1732.00        Matches:        338
Percent Similarity:     86.01%         Conservative:   0
Best Local Similarity:  86.01%         Mismatches:     0
Query Match:            96.33%         Indels:         55
DB:                     17             Gaps:           2

US-09-980-881A-3 (1-338) x AAT11671 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAA 1283

RESULT 6
AAT62846
ID   AAT62846 standard; DNA; 1749 BP.
XX
AC   AAT62846;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
```

```
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human PCPB
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
FT   misc_binding    134..177
FT                   /*tag=  d
FT                   /bound_moeity= 46_bp_probe
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   P-PSDB; AAW14733.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence encodes human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:
Pred. No.:              1.08e-180      Length:          1749
Score:                  1732.00        Matches:         338
Percent Similarity:     86.01%         Conservative:    0
Best Local Similarity:  86.01%         Mismatches:      0
Query Match:            96.33%         Indels:          55
DB:                     18             Gaps:            2

US-09-980-881A-3 (1-338) x AAT62846 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 |||||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 7
AAQ41001
ID   AAQ41001 standard; cDNA; 1749 BP.
XX
```

```
AC   AAQ41001;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B gene.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     41..106
FT                   /*tag=  a
FT   mat_peptide     107..1312
FT                   /*tag=  b
FT                   /product= PCPB
FT   misc_feature    134..177
FT                   /*tag=  c
FT                   /function= probe
FT                   /note= "used to obtain full-length clones"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   P-PSDB; AAR36273.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Disclosure; Fig 4; 40pp; English.
XX
CC   Human plasma carboxypeptidase B was isolated from human plasma and
CC   partially sequenced. Oligonucleotide primers were designed based on
CC   the partial amino acid sequences. The primers were used in a PCR
CC   amplification to identify cDNA encoding PCPB from a human liver cDNA
CC   library. The PCR product was capable of encoding the first 37 amino
CC   acids of PCPB; a 46mer probe was used to obtain the full-length
CC   sequence which, although disclosed in the specification, is not
CC   claimed. PCPB inhibits the enzymatic conversion by tPA of
CC   plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;

Alignment Scores:
Pred. No.:              3.85e-180      Length:          1749
Score:                  1727.00        Matches:         337
Percent Similarity:     85.75%         Conservative:    0
Best Local Similarity:  85.75%         Mismatches:      1
Query Match:            96.05%         Indels:          55
DB:                     14             Gaps:            2

US-09-980-881A-3 (1-338) x AAQ41001 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 8
AAF18005
```

```
ID   AAF18005 standard; DNA; 1400 BP.
XX
AC   AAF18005;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   P-PSDB; AAB58129.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 1; Page 507; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

Alignment Scores:
Pred. No.:              2.88e-89      Length:          1400
Score:                  902.00        Matches:         181
Percent Similarity:     76.05%        Conservative:    0
Best Local Similarity:  76.05%        Mismatches:      1
Query Match:            50.17%        Indels:          56
DB:                     21            Gaps:            2

US-09-980-881A-3 (1-338) x AAF18005 (1-1400)

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 175
              ||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db         10 GGAATCCATGCCAGANAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATA 69

Qy        175 ------------------------------------------------------------ 175

Db         70 ACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTAT 129

Qy        176 ----------------------------------------------AsnArgMetTr 179
                                                             ||||||||||
Db        130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACTACTCATGGAAAAAGAATCGAATGTG 189

Qy        179 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAA 249

Qy        199 nPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTy 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309

Qy        219 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAG 369

Qy        239 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 AAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTT 429

Qy        259 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 TCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAG 489

Qy        279 rGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 TGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGG 549

Qy        299 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIl 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609

Qy        319 eLysTyrSerPhe---------------------------------------------- 323
              |||||||||||
Db        610 CAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGA 669

Qy        324 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
                   ||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 719

RESULT 9
AAQ90600
ID   AAQ90600 standard; DNA; 1215 BP.
XX
AC   AAQ90600;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB; ds.
XX
OS   Sus scrofa.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1215
FT                   /*tag=  a
FT                   /product= Porcine procarboxypeptidase B.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
```

```
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   P-PSDB; AAR75131.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Claim 2; Page 20-21; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

Alignment Scores:
Pred. No.:                3.6e-61        Length:          1215
Score:                    646.50         Matches:         137
Percent Similarity:       54.57%         Conservative:    60
Best Local Similarity:    37.95%         Mismatches:      121
Query Match:              35.96%         Indels:          43
DB:                       16             Gaps:            6

US-09-980-881A-3 (1-338) x AAQ90600 (1-1215)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::     :::|||                 :::   :::|||
Db         28 TTCGAAGGGGAGAAGGTGTTCCGTGTCAATGTTGAAGATGAAAATGACATCAGCTTACTC 87

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
                 |||   :::|||   :::|||      |||:::|||   :::          |||
Db         88 CATGAGTTGGCCAGCACCAGGCAGATTGACTTCTGGAAACCAGATTCTGTCACACAAATC 147

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                     |||   |||   |||   |||   |||:::     |||:::
Db        148 AAACCTCACAGTACAGTTGACTTCCGTGTGAAAGCAGAAGATATTTTGGCTGTGGAAGAC 207

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
                |||      :::   :::         ||||||:::   ::::::      ::::::
Db        208 TTTCTGGAGCAGAATGAACTACAATATGAGGTACTCATAAACAACCTGAGATCTGTGCTC 267

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||      :::               |||   :::         ||||||:::
Db        268 GAGGCTCAGTTTGACAGCAGAGTC---------CGTACAACTGGACACAGTTATGAGAAG 318

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||::::      |||   :::|||   :::   :::|||      :::||||||
Db        319 TACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAAAATCCAGAC 378

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ::::::::::::      ||||||::::::|||           :::|||:::||||||
Db        379 CTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATACCTCCTCAAG 438

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||            |||   ||||||::::::|||||||||   ||||||
Db        439 GTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 495

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 173
              ||||||||||||||   |||||||||   ||||||:::
Db        496 AGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTTCTCACCTAT 555

Qy        173 ---------- 173

Db        556 GGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTCTTGCCTGTG 615

Qy        174 ----------GlyHis------------------AsnArgMetTrpArgLysAsnArg 183
                       |||:::                  ||||||||||||||||    |||
Db        616 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCGC 675

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||       |||         ||||||||||||||   ||||||||||
Db        676 TCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 732

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                ||||||      |||||||||:::      |||   |||||||||||||||
Db        733 GGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGCT 792

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        793 GCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTCC 852

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||:::::::::||||||||||||||||   |||::::::||||||||||||
Db        853 ATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTAT 912

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                              :::::::::   ||||||:::   :::|||      ||||||:::
Db        913 GATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAAA 972

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                 :::      |||         |||:::||||||:::|||   |||:::   |||:::
Db        973 GAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAATC 1029

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   |||   ||||||   |||||||||   ||||||   ||||||||||||||||||
Db       1030 TATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTTC 1089

Qy        324 Thr 324
              |||
Db       1090 ACC 1092
```

```
RESULT 10
AAT42494
ID   AAT42494 standard; DNA; 1263 BP.
XX
AC   AAT42494;
XX
DT   12-FEB-1997  (first entry)
XX
DE   Human pancreatic carboxypeptidase B encoding sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1248
FT                   /*tag=  a
FT                   /product= preproenzyme
FT   sig_peptide     1..39
FT                   /*tag=  b
FT                   /note= "partial enzyme secretion leader sequence
FT                        (pre-sequence); residues -108 to -96"
FT   sig_peptide     40..324
FT                   /*tag=  c
FT                   /note= "pro-sequence; residues -95 to -1"
FT   mat_peptide     385..1245
FT                   /*tag=  d
FT                   /note= "mature enzyme; encodes residues +1 to +307"
XX
```

```
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JP, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06172.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 14; Page 131; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence encodes the native prepro-HCPB (see
CC   AAW06172) contained in vector pICI1698.
XX
SQ   Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;

Alignment Scores:
Pred. No.:              3.75e-60        Length:         1263
Score:                  637.50          Matches:        130
Percent Similarity:     54.85%          Conservative:   68
Best Local Similarity:  36.01%          Mismatches:     120
Query Match:            35.46%          Indels:         43
DB:                     17              Gaps:           6

US-09-980-881A-3 (1-338) x AAT42494 (1-1263)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::     :::|||                 :::   ::::::
Db         58 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 117

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
             :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db        118 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 177

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                       |||   |||   |||   |||        |||:::
Db        178 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 237

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
             |||     :::   :::         ||||||:::::::::::   :::::::::
Db        238 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 297

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
             :::   |||       :::               ||||||:::        ||||||:::
Db        298 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 348

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||:::           |||   :::|||   :::   :::        :::|||
Db        349 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 408

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              :::::::::::     ||||||::::::||||||        :::|||:::||||||
Db        409 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 468

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||            |||   ||||||::::::||||||||   ||||||
Db        469 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 525

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 173
              ||||||||||||||||||||||||||   ||||||:::
Db        526 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 585

Qy        173 ---------- 173

Db        586 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 645

Qy        174 ----------GlyHis----------------AsnArgMetTrpArgLysAsnArg 183
                         |||:::                   :::|||   ||||||||   |||
Db        646 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 705

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::      :::   ||||||||||||||   ||||||||||||
Db        706 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 762

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                  |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db        763 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 822

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        823 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 882

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||:::::::::|||||||||||||||||   :::::::::||||||||||||
Db        883 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 942

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                              :::::::::   ||||||:::   :::|||         |||:::
Db        943 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1002

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                  :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db       1003 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1059

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   |||   ||||||   |||||||||   ||||||   ||||||:::|||||||||
Db       1060 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1119

Qy        324 Thr 324
              |||
Db       1120 ACC 1122

RESULT 11
AAV41795
ID   AAV41795 standard; DNA; 1263 BP.
XX
AC   AAV41795;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Human pancreatic carboxypeptidase nucleotide sequence.
XX
KW   ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW   prodrug therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1248
FT                   /*tag=  a
FT                   /product= "Pancreatic carboxypeptidase B"
FT                   /note= "No start codon given"
XX
PN   WO9835988-A1.
XX
PD   20-AUG-1998.
XX
```

```
PF   10-FEB-1998;   98WO-GB00415.
XX
PR   29-OCT-1997;   97GB-0022727.
PR   14-FEB-1997;   97GB-0003104.
PR   18-OCT-1997;   97GB-0022003.
XX
PA   (ZENE ) ZENEPA   (ZENE ) ZENECA LTD.
XX
PI   Edge MD;
XX
DR   WPI; 1998-467168/40.
DR   P-PSDB; AAW74476.
XX
PT   New modified pro-domain of carboxy-peptidase B - enhances expression
PT   of co-expressed proteins for production of recombinant
PT   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT   enzyme prodrug therapy
XX
PS   Example 1; Page 52-53; 83pp; English.
XX
CC   The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC   modified pro-domain of CPB on a separate gene to enhance recombinant
CC   expression.  This process can be used to produce recombinant CPB in
CC   eukaryotic cells, or fusions of CPB with antibody chains.  CPB is used in
CC   insulin production and protein sequencing, while its fusions with
CC   antibodies are useful in antibody-directed enzyme prodrug therapy.  The
CC   modified pro-domain provide increased yields of recombinant CPB, possibly
CC   by protecting the C-terminus against enzymatic degradation or by
CC   increasing intracellular trafficking.
XX
SQ   Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

Alignment Scores:
Pred. No.:                3.75e-60      Length:          1263
Score:                    637.50        Matches:         130
Percent Similarity:       54.85%        Conservative:    68
Best Local Similarity:    36.01%        Mismatches:      120
Query Match:              35.46%        Indels:          43
DB:                       19            Gaps:            6

US-09-980-881A-3 (1-338) x AAV41795 (1-1263)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||::   :::|||                          :::  ::::::
Db         58 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 117

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
             :::  |||  :::|||  :::|||    |||:::|||  :::          |||
Db        118 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 177

Qy         41 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60
                    |||  |||  |||  |||  |||         |||:::
Db        178 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 237

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
               |||      :::  :::         |||||:::::::::::  :::::::::
Db        238 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 297

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
             :::  |||     :::          |||||:::          ||||||:::
Db        298 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 348

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||::           |||  :::|||  :::  :::        :::|||
Db        349 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 408

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
             ::::::::::      ||||||:::::|||||        :::|||:::||||||
Db        409 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 468

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||  ||||||          |||  ||||||:::::|||||||  ||||||
Db        469 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 525

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 173
              ||||||||||||||||||||||||||  ||||||::
Db        526 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 585

Qy        173 ---------- 173

Db        586 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 645

Qy        174 -----------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 183
                        |||:::               :::|||  |||||||||  |||
Db        646 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 705

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::      :::  ||||||||||||||  ||||||||||||
Db        706 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 762

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                |||||||||  |||||||||  :::  |||  ||||||||||||||
Db        763 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 822

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||   |||  ||||||:::|||  |||:::|||      ::::::
Db        823 GCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 882

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||:::::::::||||||||||||||  :::::::::||||||||||||
Db        883 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 942

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                          :::::::::  ||||||:::  :::|||        |||:::
Db        943 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1002

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db       1003 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1059

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   |||  ||||||  ||||||||  ||||||  ||||||:::|||||||||
Db       1060 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1119

Qy        324 Thr 324
              |||
Db       1120 ACC 1122

RESULT 12
AAT42506
ID   AAT42506 standard; DNA; 1284 BP.
XX
AC   AAT42506;
XX
DT   17-FEB-1997  (first entry)
XX
DE   ProHCPB gene with PelB leader sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1275
FT                   /*tag=  a
FT   sig_peptide     1..66
FT                   /*tag=  b
FT                   /note= "PelB leader sequence"
FT   sig_peptide     67..351
FT                   /*tag=  c
FT                   /note= "pro sequence"
FT   mat_peptide     352..1272
```

```
FT                   /*tag=  d
FT                   /note= "mature HCPB sequence"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06175.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 18; Page 140-142; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence is a cloned pro-HCPB gene contained in
CC   plasmid pICI1738 and which can be expressed in E. coli.
XX
SQ   Sequence 1284 BP; 355 A; 315 C; 299 G; 315 T; 0 other;

Alignment Scores:
Pred. No.:              3.84e-60        Length:         1284
Score:                  637.50          Matches:        130
Percent Similarity:     54.85%          Conservative:   68
Best Local Similarity:  36.01%          Mismatches:     120
Query Match:            35.46%          Indels:         43
DB:                     17              Gaps:           6

US-09-980-881A-3 (1-338) x AAT42506 (1-1284)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::     :::|||                 :::   ::::::
Db         85 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 144

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db        145 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 204

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                        |||   |||   |||   |||   |||         |||:::
Db        205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 264

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
                 |||     :::   :::         ||||||::::::::::::   :::::::::
Db        265 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||       :::                 ||||||:::         ||||||:::
Db        325 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 375

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||:::           |||   :::|||   :::   :::         :::|||
Db        376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 435

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
```

```
                  ::::::::::::      ||||||::::::||||||         :::|||:::||||||
Db        436 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 495

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||             |||   ||||||::::::||||||||   ||||||
Db        496 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173
              ||||||||||||||||||||||||||||   ||||||:::
Db        553 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 612

Qy        173 ------------------------------------------------------------ 173

Db        613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 672

Qy        174 ------------GlyHis------------------AsnArgMetTrpArgLysAsnArg 183
                        |||:::                  :::|||   |||||||||   |||
Db        673 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 732

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::       :::   ||||||||||||||   ||||||||||||
Db        733 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db        790 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db        850 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 909

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||:::::::::||||||||||||||||   :::::::::||||||||||||
Db        910 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 969

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                              :::::::::   ||||||::   :::|||         |||:::
Db        970 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              :::     :::         |||:::||||||:::|||   |||:::   |||:::
Db       1030 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   |||   ||||||   |||||||||   ||||||   ||||||:::|||||||||
Db       1087 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146

Qy        324 Thr 324
              |||
Db       1147 ACC 1149

RESULT 13
ABZ75116
ID   ABZ75116 standard; DNA; 2023 BP.
XX
AC   ABZ75116;
XX
DT   30-MAY-2003  (first entry)
XX
DE   Anti-human seminal plasma protein ScFv/hCPA fusion gene.
XX
KW   Single chain antibody; ScFv; anti-human seminal plasma protein;
KW   monoclonal antibody E4B7; mouse; murine; human; carboxypeptidase A;
KW   hCPA; fusion protein; antibody-directed enzyme prodrug therapy; ADEPT;
KW   methotrexate-alpha-peptide prodrug; prostate cancer; cytostatic;
KW   drug targeting; heavy chain variable region; VH; VL;
KW   light chain variable region; gene; ds.
XX
OS   Chimeric - Mus musculus.
OS   Chimeric - Homo sapiens.
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             7..2010
FT                   /*tag= a
FT                   /product= "Anti-human seminal plasma protein single-chain
FT                            antibody (ScFv)/human carboxypeptidase A
FT                            (hCPA) fusion protein"
FT                   /transl_except= (pos:652..654,aa:Asn)
FT                   /transl_except= (pos:745..747,aa:Cys)
FT   mat_peptide     7..741
FT                   /*tag= b
FT                   /note= "Encodes ScFv portion of fusion protein"
FT   misc_feature    742..759
FT                   /*tag= c
FT                   /note= "Linker DNA"
FT   mat_peptide     760..2007
FT                   /*tag= d
FT                   /note= "Encodes hCPA portion of fusion protein"
XX
PN   WO2002100431-A1.
XX
PD   19-DEC-2002.
XX
PF   08-JUN-2001; 2001WO-CN00924.
XX
PR   08-JUN-2001; 2001WO-CN00924.
XX
PA   (UYTA-) UNIV TANGDU HOSPITAL FORTH MILITARY MEDI.
PA   (HAOX/) HAO X.
XX
PI   Hao X;
XX
DR   WPI; 2003-156924/15.
DR   P-PSDB; ABP97414.
XX
PT   Pharmaceutical kits for use in the treatment of prostate cancer,
PT   comprise anti-human seminal plasma protein single-chain antibody/human
PT   carboxypeptidase fusion protein and a prodrug of
PT   methotrexate-alpha-peptides -
XX
PS   Claim 4; Fig 3; 29pp; Chinese.
XX
CC   The invention relates to a pharmaceutical kit for antibody-directed
CC   enzyme prodrug therapy (ADEPT) to treat prostate cancer. The kit
CC   comprises several respective containers for holding anti-human seminal
CC   plasma protein single-chain antibody (ScFv)/human carboxypeptidase A
CC   (hCPA) fusion protein (ABP97414), a prodrug of methotrexate-alpha-peptide
CC   and a pharmaceutically-acceptable auxiliary for medication. The invention
CC   also encompasses the anti-human seminal plasma protein ScFv/hCPA fusion
CC   protein, and the nucleic acid encoding it (ABZ75116). This was
CC   constructed by linking DNA encoding the ScFv (itself generated by linking
CC   DNAs encoding the heavy and light chain variable regions (VH and VL) of
CC   the murine anti-human seminal plasma protein monoclonal antibody E4B7)
CC   to DNA encoding human carboxypeptidase A via a linker. The ScFv portion
CC   of the fusion protein enables it to be specifically targeted to prostate
CC   cancer cells, where the hCPA portion can then activate the prodrug which
CC   in turn has cytostatic activity towards the cancer cells. The invention
CC   permits the fusion protein to be easily produced in large quantities for
CC   use in ADEPT for prostate cancer treatment. The present sequence
CC   represents DNA encoding the ScFv/hCPA fusion protein of the invention.
XX
SQ   Sequence 2023 BP; 581 A; 465 C; 467 G; 510 T; 0 other;

Alignment Scores:
Pred. No.:                3.88e-58      Length:          2023
Score:                    622.00        Matches:         128
Percent Similarity:       55.12%        Conservative:    71
Best Local Similarity:    35.46%        Mismatches:      118
Query Match:              34.59%        Indels:          44
DB:                       25            Gaps:            7

US-09-980-881A-3 (1-338) x ABZ75116 (1-2023)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||         :::|||             |||:::      :::|||      ::::::

Db        817 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 876

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ::::::|||      |||   |||:::      |||   |||                :::

Db        877 AAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTA 936

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                           |||   |||   |||:::      :::         :::::::::

Db        937 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 996

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||:::   :::   :::         :::|||:::   |||:::::::::   |||

Db        997 GCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 1056

Qy         81 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 99
              ::::::|||           :::   :::      |||   |||         |||

Db       1057 GAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCA 1107

Qy        100 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 119
              :::|||::::::      :::|||   :::|||   |||   :::   :::::::::|||

Db       1108 AAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCT 1167

Qy        120 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 139
              :::|||:::::::::|||   |||||||||:::   |||      |||||||||||||||

Db       1168 GAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTG 1227

Qy        140 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 159
              |||:::   |||::::::         :::   ||||||:::::|||||||||||||||

Db       1228 AAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCAC 1284

Qy        160 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 173
              ||||||||||||:::|||||||||||||||   ||||||:::

Db       1285 GCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACT 1344

Qy        173 ------------------------------------------------------------ 173

Db       1345 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 1404

Qy        174 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 182
                            |||:::                 ||||||||||||||||||

Db       1405 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 1464

Qy        183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
              ||||||          |||:::   ||||||||||||||||||||||||||||   :::

Db       1465 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 1521

Qy        203 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 222
                    |||               ::::::      |||::::::   |||   |||

Db       1522 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 1581

Qy        223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
                 ||||||||||||   |||   |||||||||   :::|||::|||   ::::::|||

Db       1582 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 1641

Qy        243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 262
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||

Db       1642 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 1701

Qy        263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
                 ||||||               :::||||||:::|||:::   ||||||

Db       1702 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1761

Qy        283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
                    :::      |||          |||||||||   :::|||            |||

Db       1762 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1818

Qy        303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
              :::|||      |||          ||||||   ||||||||||||||||:::::

Db       1819 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1878

Qy        323 Phe 323
              |||

Db       1879 TTT 1881
```

```
RESULT 14
ABV77254
ID   ABV77254 standard; DNA; 1302 BP.
XX
AC   ABV77254;
XX
DT   28-MAR-2003  (first entry)
XX
DE   Nucleotide sequence of human zinc carboxypeptidase B-like protein.
XX
KW   Human; zinc carboxypeptidase B-like protein; cancer; CNS disorder;
KW   chronic obstructive pulmonary disease; COPD; haematological disorder;
KW   gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1302
FT                   /*tag= a
FT                   /product= "zinc carboxypeptidase B-like protein"
FT                   /note= "no termination codon given"
XX
PN   WO200295018-A1.
XX
PD   28-NOV-2002.
XX
PF   16-MAY-2002; 2002WO-EP05392.
XX
PR   22-MAY-2001; 2001US-292329P.
XX
PA   (FARB ) BAYER AG.
XX
PI   Smolyar A;
XX
DR   WPI; 2003-120801/11.
DR   P-PSDB; ABB99687.
XX
PT   New polynucleotide encoding a zinc carboxypeptidase B-like protein
PT   polypeptide, useful for treating diseases associated with
PT   carboxypeptidase dysfunction, e.g. cancer, CNS disorder, COPD or
PT   haematological disorder  -
XX
PS   Claim 1; Fig 1; 113pp; English.
XX
CC   The present sequence encodes a human zinc carboxypeptidase B-like
CC   protein. The zinc carboxypeptidase B-like protein is used for treating
CC   diseases, such as cancer, a CNS disorder, chronic obstructive pulmonary
CC   disease (COPD) or haematological disorders. The zinc carboxypeptidase
CC   B-like protein may also be used to identify compounds which may act
CC   as activators or inhibitors at the enzyme's active site, to raise
CC   specific antibodies which can block the enzyme and effectively reduce
CC   its activity, as a bait protein in a two-hybrid or three-hybrid assay
CC   to identify other proteins which bind to or interact with the human
CC   zinc carboxypeptidase-like protein polypeptide and modulate its
CC   activity, and for immunization of mammals.
XX
SQ   Sequence 1302 BP; 388 A; 275 C; 283 G; 356 T; 0 other;

Alignment Scores:
Pred. No.:              3.8e-58       Length:          1302
Score:                  619.50        Matches:         136
Percent Similarity:     51.15%        Conservative:    64
Best Local Similarity:  34.78%        Mismatches:      127
Query Match:            34.45%        Indels:          64
DB:                     25            Gaps:            8

US-09-980-881A-3 (1-338) x ABV77254 (1-1302)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              :::         :::|||:::      :::|||:::|||      :::          |||
Db        109 TATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTG 168

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
               :::    ::::::          ::::::   ||||||||||||   :::         :::
Db        169 AAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCCAGCAGTATCTCCTATGTA 228

Qy         41 ValLys------LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 58
                 :::          ||||||          |||   |||             :::
Db        229 TCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCCGAGCC------CTG 282

Qy         59 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 78
              |||    |||       :::   |||          |||||:::   |||::::::
Db        283 TTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAGAAGATCTTCAGAAA 342

Qy         79 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer------ 96
                 :::::::::       |||      |||       |||    |||    |||
Db        343 ACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGATCCCTCTCTGGATAT 402

Qy         97 TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 116
              ||||||   ||||||||||||   ||||||   :::|||:::       :::   :::
Db        403 AATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAA 462

Qy        117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 136
              |||       ::::::             |||||    |||:::|||           |||
Db        463 ACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTT 522

Qy        137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCys 156
              ::::::|||||:::   |||:::          |||   |||:::||||||||||||
Db        523 TTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGT 579

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle--------- 173
              ||||||||||||||||||||||||   ||||||||||||   ||||||:::
Db        580 GGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCT 639

Qy        173 ------------------------------------------------------------ 173

Db        640 CTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTAT 699

Qy        174 ----------------------------------------GlyHisAsnArgMetTrp 179
                                                       ::::::|||   |||
Db        700 ATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGG 759

Qy        180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199
              ||||||   ||||||       :::          |||   |||   |||   |||||||||
Db        760 AGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAAC 819

Qy        200 PheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyr 219
              :::      |||   ||||||:::||||||||||||          |||   :::||||||
Db        820 TGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATAC 876

Qy        220 CysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArg 239
              ||||||   :::||||||||||||||||||||||||||||||||:::|||||||||:::
Db        877 TGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAA 936

Qy        240 AsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhe 259
              :::         |||:::||||||:::|||    |||:::|||:::|||   :::::::::
Db        937 CACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTAT 996

Qy        260 ProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSer 279
              ||||||||||||       :::          :::          :::          |||
Db        997 CCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTAT 1056

Qy        280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 299
              :::||||||   |||::::::                ||||||   :::|||
Db       1057 AAAGCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCC 1113

Qy        300 SerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIle 319
              |||   |||||||||::::::   |||          ||||||   |||      ||||||
Db       1114 TCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATA 1173

Qy        320 LysTyrSerPhe----------------------------------------------- 323
                 |||:::|||
Db       1174 CCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAG 1233

Qy        324 ----ThrSerAsnProProValGluLysLeu 332
```

```
                  :::|||||||||||||||||:::||||||
Db       1234 ATGCTCATCAAACCCACCTGTACAGAAACTA 1264

RESULT 15
AAD03837
ID   AAD03837 standard; cDNA; 1311 BP.
XX
AC   AAD03837;
XX
DT   19-JUN-2001  (first entry)
XX
DE   Human carboxypeptidase-related protein #5 cDNA.
XX
KW   Human; carboxypeptidase-related protein; protease; gene therapy;
KW   drug screening; mental; biological disorder; medical disorder; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1311
FT                   /*tag= a
FT                   /product= "Human carboxypeptidase-related protein #5"
FT                   /note= "CDS does not include stop codon"
FT                   /partial
XX
PN   WO200123588-A1.
XX
PD   05-APR-2001.
XX
PF   29-SEP-2000; 2000WO-US26876.
XX
PR   29-SEP-1999;   99US-0156685.
XX
PA   (LEXI-) LEXICON GENETICS INC.
XX
PI   Donoho G,  Turner CA,  Nehls M,  Friedrich G,  Zambrowicz B,  Sands AT;
XX
DR   WPI; 2001-266171/27.
DR   P-PSDB; AAE00502.
XX
PT   New isolated human carboxypeptidase polynucleotide useful for
PT   generating antibodies, as reagents in diagnostic assays and for
PT   screening for compounds useful for treating mental, biological or
PT   medical diseases -
XX
PS   Claim 1; Page 30; 36pp; English.
XX
CC   The present sequence is a cDNA encoding novel human protein (NHP),
CC   carboxypeptidase-related protein. The carboxypeptidase-related
CC   proteins share structural similarity with animal proteases, and
CC   especially carboxypeptidase B or carboxypeptidase A. The
CC   carboxypeptidase-related protein cDNA is useful for the detection of
CC   mutant human carboxypeptidases, or inappropriately expressed human
CC   carboxypeptidases for the diagnosis of disease, and also as a
CC   therapeutic. It is also useful in drug screening, for generation of
CC   antibodies, as reagents in diagnostic assays, for the identification of
CC   other cellular gene products related to human carboxypeptidases, and as
CC   reagents in assays for screening for compounds that are useful for
CC   treating mental, biological or medical disorders and diseases. Nucleotide
CC   constructs encoding functional NHPs can also be used in gene therapy.
XX
SQ   Sequence 1311 BP; 391 A; 278 C; 284 G; 358 T; 0 other;

Alignment Scores:
Pred. No.:              3.84e-58        Length:         1311
Score:                  619.50          Matches:        136
Percent Similarity:     51.15%          Conservative:   64
Best Local Similarity:  34.78%          Mismatches:     127
Query Match:            34.45%          Indels:         64
DB:                     22              Gaps:           8

US-09-980-881A-3 (1-338) x AAD03837 (1-1311)
```

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              :::         :::|||:::      :::|||:::|||      :::          |||
Db        109 TATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTG 168

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   ::::::       ::::::   ||||||||||||   :::          :::
Db        169 AAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCCAGCAGTATCTCCTATGTA 228

Qy         41 ValLys------LysLysGlnValHisPheValAsnAlaSerAspValAspAsnVal 58
                 :::          ||||||         |||   |||           :::
Db        229 TCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCCGAGCC------CTG 282

Qy         59 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 78
              |||   |||       :::   |||          ||||||:::   |||::::::
Db        283 TTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAGAAGATCTTCAGAAA 342

Qy         79 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer------ 96
                :::::::::          |||       |||          |||   |||   |||
Db        343 ACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGATCCCTCTCTGGATAT 402

Qy         97 TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 116
              ||||||   ||||||||||||   ||||||   :::|||:::       :::   :::
Db        403 AATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAA 462

Qy        117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 136
              |||       ::::::              ||||||   |||:::|||          |||
Db        463 ACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTT 522

Qy        137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCys 156
              ::::::||||||:::   |||:::              |||   |||:::||||||||||||
Db        523 TTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGT 579

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle--------- 173
              ||||||||||||||||||||||||   ||||||||||||   ||||||:::
Db        580 GGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCT 639

Qy        173 ------------------------------------------------------------ 173

Db        640 CTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTAT 699

Qy        174 ----------------------------------------GlyHisAsnArgMetTrp 179
                                                       ::::::|||   |||
Db        700 ATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGG 759

Qy        180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199
              ||||||   ||||||      :::         |||   |||   |||   |||||||||
Db        760 AGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAAC 819

Qy        200 PheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyr 219
              :::      |||   ||||||:::||||||||||||||          |||   :::||||||
Db        820 TGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATAC 876

Qy        220 CysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArg 239
              ||||||   :::||||||||||||||||||||||||||||||||:::|||||||||:::
Db        877 TGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAA 936

Qy        240 AsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhe 259
              :::         |||:::||||||:::|||   |||:::|||:::|||   :::::::::
Db        937 CACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTAT 996

Qy        260 ProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSer 279
              ||||||||||||      :::         :::         :::          |||
Db        997 CCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTAT 1056

Qy        280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 299
              :::||||||   |||::::::                ||||||   :::|||
Db       1057 AAAGCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCC 1113

Qy        300 SerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIle 319
              |||   |||||||||::::::   |||         ||||||   |||      ||||||
Db       1114 TCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATA 1173

Qy        320 LysTyrSerPhe----------------------------------------------- 323
```

```
                |||:::|||
Db       1174 CCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAG 1233

Qy        324 ----ThrSerAsnProProValGluLysLeu 332
                :::||||||||||||||||:::||||||
Db       1234 ATGCTCATCAAACCCACCTGTACAGAAACTA 1264
```

Search completed: January 1, 2004, 20:09:23
Job time : 391.048 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:51:12 ; Search time 155.24 Seconds
(without alignments)
961.011 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135917_26894/app_query.fasta_1.1628
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_115_@runat_31122003_135917_26894 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1749.5	97.3	1625	4	US-09-813-133A-1	Sequence 1, Appli
2	1732	96.3	1272	2	US-08-869-057-1	Sequence 1, Appli
3	1732	96.3	1749	1	US-07-649-591B-2	Sequence 2, Appli
4	1732	96.3	1749	1	US-08-277-540-2	Sequence 2, Appli
5	1732	96.3	1749	1	US-08-430-787A-2	Sequence 2, Appli
6	646.5	36.0	1215	1	US-08-696-139-1	Sequence 1, Appli
7	637.5	35.5	1263	2	US-08-860-882A-56	Sequence 56, Appl
8	637.5	35.5	1263	4	US-09-011-769A-38	Sequence 38, Appl
9	637.5	35.5	1284	2	US-08-860-882A-71	Sequence 71, Appl
10	637.5	35.5	1284	4	US-09-011-769A-55	Sequence 55, Appl
11	619.5	34.5	1311	4	US-09-675-305-9	Sequence 9, Appli
12	619.5	34.5	2154	3	US-09-171-945-124	Sequence 124, App
13	610	33.9	927	2	US-08-782-760-5	Sequence 5, Appli
14	610	33.9	927	5	PCT-US96-00995-5	Sequence 5, Appli
15	582	32.4	921	1	US-08-696-139-3	Sequence 3, Appli
16	571.5	31.8	2128	4	US-09-675-305-13	Sequence 13, Appl
17	567	31.5	999	2	US-08-860-882A-67	Sequence 67, Appl
18	567	31.5	999	4	US-09-011-769A-50	Sequence 50, Appl
19	567	31.5	1053	2	US-08-860-882A-64	Sequence 64, Appl
20	567	31.5	1053	4	US-09-011-769A-46	Sequence 46, Appl
21	560	31.1	1053	4	US-09-463-451-27	Sequence 27, Appl
c 22	560	31.1	1053	4	US-09-463-451-28	Sequence 28, Appl
23	560	31.1	1059	2	US-08-860-882A-74	Sequence 74, Appl
24	560	31.1	1059	4	US-09-011-769A-59	Sequence 59, Appl
25	559	31.1	1059	2	US-08-860-882A-77	Sequence 77, Appl
26	559	31.1	1059	4	US-09-011-769A-63	Sequence 63, Appl
27	554	30.8	1200	4	US-09-710-099-7	Sequence 7, Appli
28	549	30.5	1870	3	US-09-171-945-112	Sequence 112, App
29	534.5	29.7	1050	4	US-09-675-305-11	Sequence 11, Appl
30	527.5	29.3	1311	4	US-09-710-099-5	Sequence 5, Appli
31	516.5	28.7	1251	3	US-08-640-906-3	Sequence 3, Appli
32	516.5	28.7	1251	4	US-09-395-936-3	Sequence 3, Appli
33	491.5	27.3	1257	3	US-08-640-906-1	Sequence 1, Appli
34	491.5	27.3	1257	4	US-09-395-936-1	Sequence 1, Appli
35	467.5	26.0	945	4	US-09-710-099-3	Sequence 3, Appli
36	467.5	26.0	945	4	US-09-710-099-11	Sequence 11, Appl
37	465.5	25.9	2247	4	US-09-710-099-15	Sequence 15, Appl
38	441	24.5	1056	4	US-09-710-099-1	Sequence 1, Appli
39	441	24.5	1056	4	US-09-710-099-9	Sequence 9, Appli
40	365	20.3	55827	4	US-09-813-133A-3	Sequence 3, Appli
41	288.5	16.0	741	4	US-09-675-305-5	Sequence 5, Appli
42	210.5	11.7	591	3	US-09-331-709-2	Sequence 2, Appli
43	198.5	11.0	629	4	US-09-280-116-228	Sequence 228, App
44	169.5	9.4	494	4	US-09-669-751-63	Sequence 63, Appl
c 45	169.5	9.4	515	3	US-08-998-416-125	Sequence 125, App

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Alignment Scores:

Pred. No.:	8.02e-210	Length:	1625
Score:	1749.50	Matches:	336
Percent Similarity:	94.38%	Conservative:	0
Best Local Similarity:	94.38%	Mismatches:	2
Query Match:	97.30%	Indels:	18
DB:	4	Gaps:	1

US-09-980-881A-3 (1-338) x US-09-813-133A-1 (1-1625)

Qy 1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
Db 83 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 142

Qy 21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

```
Db        143 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 202

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 262

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 322

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 382

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 442

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 502

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Db        503 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 562

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsnArgMetTrpArg 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATAATCGAATGTGGAGA 622

Qy        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        623 AAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTT 682

Qy        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        683 GCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGT 742

Qy        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 GGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAT 802

Qy        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 ATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCA 862

Qy        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAA 922

Qy        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
              |||||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db        923 GCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGGCCATGGCTCA 982

Qy        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 GAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAA 1042

Qy        321 TyrSerPhe---------------------------------------------------- 323
              |||||||||
Db       1043 TATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGT 1102

Qy        324 -ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
               |||||||||||||||||||||||||||||||||||||||||||||
Db       1103 TACATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAAA 1148

RESULT 2
US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1272 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    PUBLICATION INFORMATION:
;      AUTHORS:  Eaton, Dan L.
;      AUTHORS:  Malloy, Beth E.
;      AUTHORS:  Tsai, Siao P
;      AUTHORS:  Henzel, William
;      AUTHORS:  Drayna, Dennis
;      TITLE:  Isolation, Molecular Cloning, and Partial
;      TITLE:  Characterization of a No. 5985562el Carboxypeptidase B
;      TITLE:  from Human Plasma
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  266
;      ISSUE:  32
;      PAGES:  21833-21838
;      DATE:  No. 5985562 15-1991
US-08-869-057-1

Alignment Scores:
Pred. No.:                8.45e-208       Length:          1272
Score:                    1732.00         Matches:         338
Percent Similarity:       86.01%          Conservative:    0
Best Local Similarity:    86.01%          Mismatches:      0
Query Match:              96.33%          Indels:          55
DB:                       2               Gaps:            2

US-09-980-881A-3 (1-338) x US-08-869-057-1 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        187 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 666

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||
Db        667 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ------------------------------------------------ThrSerAs 326
                                                               ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              |||||||||||||||||||||||||||||||||||
Db       1207 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 3
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  hybridization probe
;      LOCATION:  133 to 178
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  potential clip site
;      LOCATION:  380 to 382
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  signal sequence
;      LOCATION:  41 to 106
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-07-649-591B-2

Alignment Scores:
Pred. No.:                1.44e-207       Length:          1749
Score:                    1732.00         Matches:         338
Percent Similarity:       86.01%          Conservative:    0
Best Local Similarity:    86.01%          Mismatches:      0
Query Match:              96.33%          Indels:          55
DB:                       1               Gaps:            2

US-09-980-881A-3 (1-338) x US-07-649-591B-2 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286
```

```
Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  ||||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              |||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 4
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-277-540-2

Alignment Scores:
Pred. No.:                1.44e-207      Length:          1749
Score:                    1732.00        Matches:         338
Percent Similarity:       86.01%         Conservative:    0
Best Local Similarity:    86.01%         Mismatches:      0
Query Match:              96.33%         Indels:          55
DB:                       1              Gaps:            2

US-09-980-881A-3 (1-338) x US-08-277-540-2 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406
```

```
Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 |||||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-787A-2

Alignment Scores:
Pred. No.:              1.44e-207       Length:         1749
Score:                  1732.00         Matches:        338
Percent Similarity:     86.01%          Conservative:   0
Best Local Similarity:  86.01%          Mismatches:     0
Query Match:            96.33%          Indels:         55
DB:                     1               Gaps:           2

US-09-980-881A-3 (1-338) x US-08-430-787A-2 (1-1749)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        107 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 166

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        167 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 226

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        227 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 286

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        287 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 346

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        347 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 406

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 466
```

```
Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 526

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        527 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 586

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              |||||||||||||||||||||||||||||||||||||||||||||
Db        587 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 646

Qy        175 ------------------------------------------------------------ 175

Db        647 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 706

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||
Db        707 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 766

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 826

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 886

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 946

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        947 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 1006

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1007 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1066

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1067 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1126

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1127 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1186

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1187 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1246

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              |||||||||||||||||||||||||||||||||||
Db       1247 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1283

RESULT 6
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1215 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1215
US-08-696-139-1

Alignment Scores:
Pred. No.:                2.96e-71        Length:          1215
Score:                    646.50          Matches:         137
Percent Similarity:       54.57%          Conservative:    60
Best Local Similarity:    37.95%          Mismatches:      121
Query Match:              35.96%          Indels:          43
DB:                       1               Gaps:            6

US-09-980-881A-3 (1-338) x US-08-696-139-1 (1-1215)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::      :::|||                                 :::   :::|||
Db         28 TTCGAAGGGGAGAAGGTGTTCCGTGTCAATGTTGAAGATGAAAATGACATCAGCTTACTC 87

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
                 |||   :::|||  :::|||      |||:::|||   :::         |||
Db         88 CATGAGTTGGCCAGCACCAGGCAGATTGACTTCTGGAAACCAGATTCTGTCACACAAATC 147

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                         |||  |||  |||  |||  |||:::      |||:::
Db        148 AAACCTCACAGTACAGTTGACTTCCGTGTGAAAGCAGAAGATATTTTGGCTGTGGAAGAC 207

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
               |||      :::  :::         ||||||:::  ::::::      ::::::
Db        208 TTTCTGGAGCAGAATGAACTACAATATGAGGTACTCATAAACAACCTGAGATCTGTGCTC 267

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||     :::          |||   :::          ||||||:::
Db        268 GAGGCTCAGTTTGACAGCAGAGTC---------CGTACAACTGGACACAGTTATGAGAAG 318

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||::::::         |||  :::|||  :::  :::|||      :::|||||
Db        319 TACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAAAATCCAGAC 378

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ::::::::::      ||||||:::::|||           :::|||:::||||||
Db        379 CTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATACCTCCTCAAG 438
```

```
Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||             |||   |||||||::::::|||||||||   ||||||
Db        439 GTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 495

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173
              ||||||||||||||   ||||||||   |||||:::
Db        496 AGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTTCTCACCTAT 555

Qy        173 ------------------------------------------------------------ 173

Db        556 GGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTCTTGCCTGTG 615

Qy        174 ------------GlyHis------------------AsnArgMetTrpArgLysAsnArg 183
                          |||:::                 |||||||||||||||||   |||
Db        616 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGAAAGACCCGC 675

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||     |||       ||||||||||||||   ||||||||||||
Db        676 TCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 732

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                  ||||||      |||||||||:::       |||   ||||||||||||||||
Db        733 GGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGTGGATCTGCT 792

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
                  |||||||||   |||   ||||||:::|||   |||:::|||   |||::::::
Db        793 GCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAACCTCTCCTCC 852

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||:::::::::||||||||||||||||   |||:::::||||||||||||
Db        853 ATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCTTATTCCTAT 912

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                          :::::::::   ||||||:::   :::|||       ||||||:::
Db        913 GATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCTGCCGTGAAA 972

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                  :::       |||       |||:::||||||:::|||   |||:::   |||:::
Db        973 GAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCTACAACAATC 1029

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   |||   ||||||   |||||||||   ||||||   ||||||||||||||||||
Db       1030 TATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAATATTCCTTC 1089

Qy        324 Thr 324
              |||
Db       1090 ACC 1092
```

```
RESULT 7
US-08-860-882A-56
; Sequence 56, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1263 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-56

Alignment Scores:
Pred. No.:                4.29e-70        Length:          1263
Score:                    637.50          Matches:         130
Percent Similarity:       54.85%          Conservative:    68
Best Local Similarity:    36.01%          Mismatches:      120
Query Match:              35.46%          Indels:          43
DB:                       2               Gaps:            6

US-09-980-881A-3 (1-338) x US-08-860-882A-56 (1-1263)
```

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::      :::|||                                 :::   ::::::
Db         58 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 117

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db        118 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 177

Qy         41 ValLysLysLysGlnValHisPheProValAsnAlaSerAspValAspAsnValLysAla 60
                              |||   |||   |||   |||   |||           |||:::
Db        178 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 237

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
                 |||      :::   :::         ||||||::::::::::::   :::::::::
Db        238 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 297

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||      :::                |||||:::           |||||:::
Db        298 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 348

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||:::          |||   :::|||   :::   :::          :::|||
Db        349 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 408

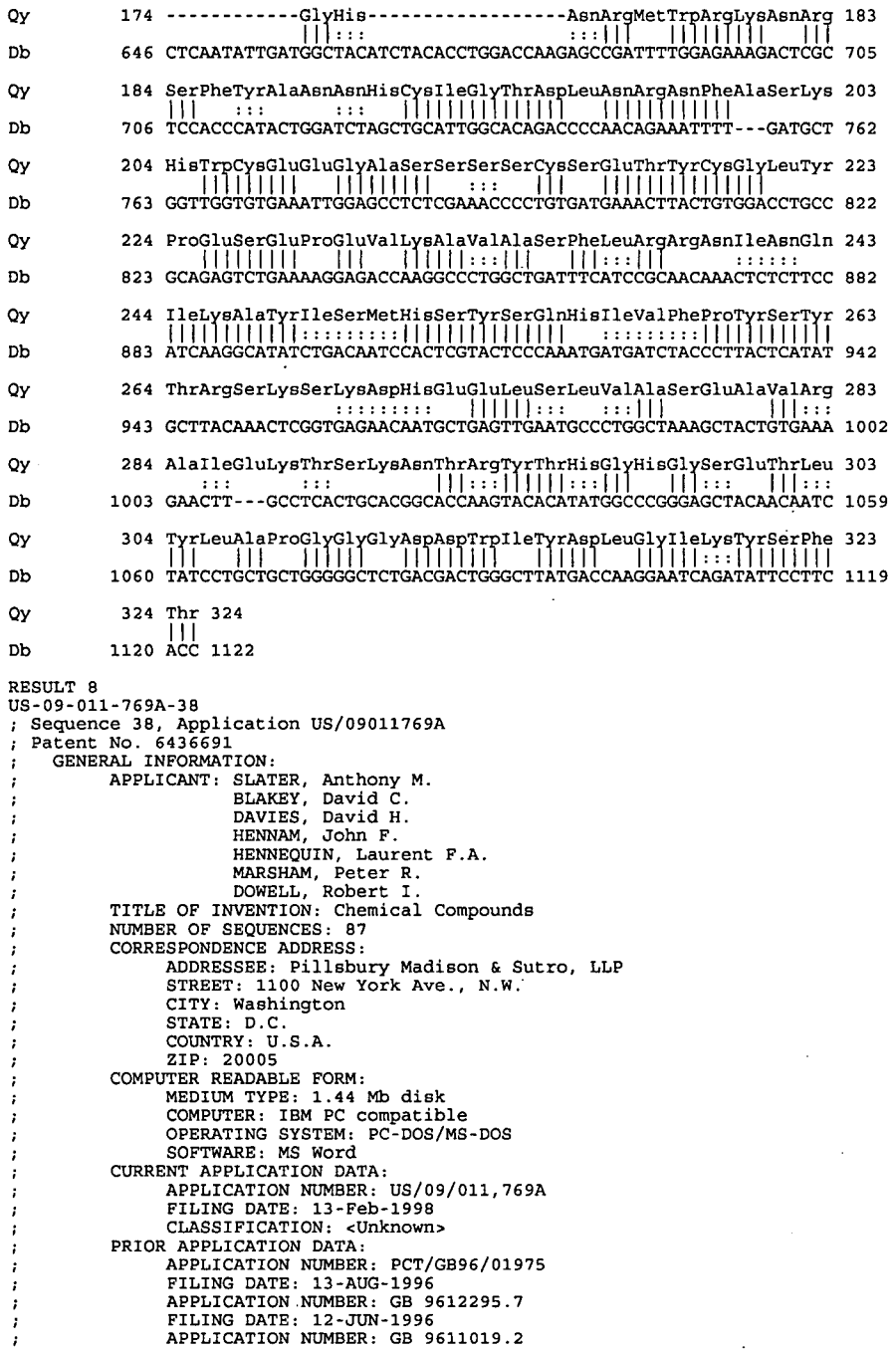
Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              :::::::::::      ||||||:::::||||||         :::|||:::||||||
Db        409 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 468

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||             |||   ||||||::::::|||||||||   ||||||
Db        469 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 525

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173
              ||||||||||||||||||||||||||||||   ||||||:::
Db        526 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 585

Qy        173 ------------------------------------------------------------ 173

Db        586 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 645
```



```
Qy       174 ------------GlyHis------------------AsnArgMetTrpArgLysAsnArg 183
                         |||:::                  :::|||   |||||||||   |||
Db       646 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 705

Qy       184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
             |||   :::      :::   ||||||||||||||||   ||||||||||||
Db       706 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 762

Qy       204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                 |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db       763 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 822

Qy       224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
                 |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db       823 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 882

Qy       244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
             ||||||||||||:::::::::||||||||||||||||   :::::::::||||||||||||
Db       883 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 942

Qy       264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                         :::::::::   ||||||:::   :::|||          |||:::
Db       943 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1002

Qy       284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                 :::      :::         |||:::|||||||:::|||   |||:::   |||:::
Db      1003 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1059

Qy       304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
             |||   |||   ||||||   |||||||||   ||||||   |||||||:::|||||||||
Db      1060 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1119

Qy       324 Thr 324
             |||
Db      1120 ACC 1122

RESULT 8
US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 38:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1263 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

Alignment Scores:
Pred. No.:                  4.29e-70        Length:          1263
Score:                      637.50          Matches:         130
Percent Similarity:         54.85%          Conservative:    68
Best Local Similarity:      36.01%          Mismatches:      120
Query Match:                35.46%          Indels:          43
DB:                         4               Gaps:            6

US-09-980-881A-3 (1-338) x US-09-011-769A-38 (1-1263)

Qy         1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
             |||:::      :::|||                                 :::   ::::::
Db        58 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 117

Qy        21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
             :::   |||   :::|||   :::|||       |||:::|||   :::         |||
Db       118 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 177

Qy        41 ValLysLysLysGlnValHisPheValAsnAlaSerAspValAspAsnValLysAla 60
                         |||   |||   |||   |||   |||           |||:::
Db       178 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 237

Qy        61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
                |||       :::   :::         ||||||:::::::::::   :::::::::
Db       238 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 297

Qy        81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
             :::   |||       :::                 ||||||:::           |||||:::
Db       298 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 348

Qy       101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
             |||:::          |||   :::|||   :::   :::           :::|||
Db       349 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 408

Qy       121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
             ::::::::::::      ||||||:::::||||||          :::|||:::||||||
Db       409 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 468

Qy       141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
             |||   ||||||              |||   ||||||::::::|||||||||   ||||||
Db       469 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 525

Qy       161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 173
             ||||||||||||||||||||||||||   ||||||:::
Db       526 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 585

Qy       173 ---------- 173

Db       586 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 645

Qy       174 ------------GlyHis------------------AsnArgMetTrpArgLysAsnArg 183
                         |||:::                  :::|||   |||||||||   |||
Db       646 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 705

Qy       184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
             |||   :::      :::   ||||||||||||||||   ||||||||||||
Db       706 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 762

Qy       204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
                 |||||||||   |||||||||   :::   |||   ||||||||||||||||
```

```
Db        763 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 822

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243

Db        823 GCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 882

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263

Db        883 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 942

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283

Db        943 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1002

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303

Db       1003 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1059

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323

Db       1060 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1119

Qy        324 Thr 324

Db       1120 ACC 1122
```

```
RESULT 9
US-08-860-882A-71
; Sequence 71, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  71:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1284 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
```

US-08-860-882A-71

Alignment Scores:

Pred. No.:	4.41e-70	Length:	1284
Score:	637.50	Matches:	130
Percent Similarity:	54.85%	Conservative:	68
Best Local Similarity:	36.01%	Mismatches:	120
Query Match:	35.46%	Indels:	43
DB:	2	Gaps:	6

US-09-980-881A-3 (1-338) x US-08-860-882A-71 (1-1284)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Db         85 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 144

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Db        145 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 204

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Db        205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 264

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Db        265 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Db        325 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 375

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Db        376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 435

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Db        436 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 495

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Db        496 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173

Db        553 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 612

Qy        173 ------------------------------------------------------------ 173

Db        613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 672

Qy        174 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 183

Db        673 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 732

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203

Db        733 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223

Db        790 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243

Db        850 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 909

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263

Db        910 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 969

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
```

```
                          :::::::::   ||||||:::   :::|||         |||:::
Db        970 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                 :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db       1030 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
               |||   |||   ||||||   ||||||||   ||||||   ||||||:::|||||||||
Db       1087 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146

Qy        324 Thr 324
               |||
Db       1147 ACC 1149

RESULT 10
US-09-011-769A-55
; Sequence 55, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 55:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 1284 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: other nucleic acid
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION: 1..1272
;        FEATURE:
;             NAME/KEY:  mat_peptide
;             LOCATION: 352..1272
;        SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-011-769A-55

Alignment Scores:
Pred. No.:              4.41e-70        Length:         1284
Score:                  637.50          Matches:        130

Percent Similarity:     54.85%          Conservative:   68
Best Local Similarity:  36.01%          Mismatches:     120
Query Match:            35.46%          Indels:         43
DB:                     4               Gaps:           6

US-09-980-881A-3 (1-338) x US-09-011-769A-55 (1-1284)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::      :::|||                  :::   ::::::
Db         85 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 144

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db        145 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 204

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                           |||  |||   |||   |||   |||         |||:::
Db        205 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 264

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
               |||      :::  :::      ||||||:::::::::::   :::::::::
Db        265 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 324

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||      :::               ||||||:::        ||||||:::
Db        325 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 375

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||:::            |||   :::|||   :::   :::         :::|||
Db        376 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 435

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ::::::::::::      ||||||:::::||||||         :::|||:::||||||
Db        436 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 495

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||            |||   ||||||:::::|||||||||   ||||||
Db        496 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 552

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173
              ||||||||||||||||||||||||||||   ||||||:::
Db        553 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 612

Qy        173 ------------------------------------------------------------ 173

Db        613 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 672

Qy        174 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 183
                         |||:::                 :::|||   |||||||||   |||
Db        673 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 732

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::       :::   ||||||||||||||||   ||||||||||
Db        733 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 789

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              |||||||||   |||||||||   :::   |||   ||||||||||||||
Db        790 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 849

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||   |||   ||||||:::|||   |||:::|||       ::::::
Db        850 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 909

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||:::::::::||||||||||||||   :::::::::||||||||||||
Db        910 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 969

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                           :::::::::   ||||||:::   :::|||         |||:::
Db        970 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1029

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                 :::      :::         |||:::||||||:::|||   |||:::   |||:::
```

```
Db       1030 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323

Db       1087 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1146

Qy        324 Thr 324

Db       1147 ACC 1149
```

RESULT 11
US-09-675-305-9
; Sequence 9, Application US/09675305
; Patent No. 6441153
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6441153el Human Carboxypeptidases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/09/675,305
; CURRENT FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: US 60/156,685
; PRIOR FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-675-305-9

Alignment Scores:

Pred. No.:	8.4e-68	Length:	1311
Score:	619.50	Matches:	136
Percent Similarity:	51.15%	Conservative:	64
Best Local Similarity:	34.78%	Mismatches:	127
Query Match:	34.45%	Indels:	64
DB:	4	Gaps:	8

US-09-980-881A-3 (1-338) x US-09-675-305-9 (1-1311)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Db        109 TATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTG 168

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Db        169 AAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCCAGCAGTATCTCCTATGTA 228

Qy         41 ValLys------LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 58

Db        229 TCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCCGAGCC------CTG 282

Qy         59 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 78

Db        283 TTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAGAAGATCTTCAGAAA 342

Qy         79 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer------ 96

Db        343 ACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGATCCCTCTCTGGATAT 402

Qy         97 TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 116

Db        403 AATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAA 462

Qy        117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 136

Db        463 ACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTT 522

Qy        137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCys 156

Db        523 TTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGT 579

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle--------- 173

Db        580 GGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCT 639

Qy        173 ---------- 173

Db        640 CTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTAT 699

Qy        174 ----------GlyHisAsnArgMetTrp 179

Db        700 ATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGG 759

Qy        180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199

Db        760 AGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAAC 819

Qy        200 PheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyr 219

Db        820 TGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATAC 876

Qy        220 CysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArg 239

Db        877 TGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAA 936

Qy        240 AsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhe 259

Db        937 CACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTAT 996

Qy        260 ProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSer 279

Db        997 CCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTAT 1056

Qy        280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 299

Db       1057 AAAGCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCC 1113

Qy        300 SerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIle 319

Db       1114 TCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATA 1173

Qy        320 LysTyrSerPhe---------- 323

Db       1174 CCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAG 1233

Qy        324 ----ThrSerAsnProProValGluLysLeu 332

Db       1234 ATGCTCATCAAACCCACCTGTACAGAAACTA 1264
```

RESULT 12
US-09-171-945-124
; Sequence 124, Application US/09171945
; Patent No. 6277599
; GENERAL INFORMATION:
; APPLICANT: Emery, Stephen
; APPLICANT: Copley, Clive Graham
; APPLICANT: Edge, Michael Derek
; TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said
; TITLE OF INVENTION: Antibody, and Their Therapeutic Use in an Adept System
; FILE REFERENCE: Monoclonal Antibody to CEA
; CURRENT APPLICATION NUMBER: US/09/171,945
; CURRENT FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: GB9703103.3
; PRIOR FILING DATE: 1997-02-14
; PRIOR APPLICATION NUMBER: GB9609405.7
; PRIOR FILING DATE: 1996-05-04
; PRIOR APPLICATION NUMBER: PCT/GB97/01165
; PRIOR FILING DATE: 1997-04-29
; NUMBER OF SEQ ID NOS: 131

```
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 124
;   LENGTH: 2154
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: humanized
US-09-171-945-124

Alignment Scores:
Pred. No.:              1.94e-67        Length:         2154
Score:                  619.50          Matches:        127
Percent Similarity:     54.29%          Conservative:   69
Best Local Similarity:  35.18%          Mismatches:     122
Query Match:            34.45%          Indels:         43
DB:                     3               Gaps:           6

US-09-980-881A-3 (1-338) x US-09-171-945-124 (1-2154)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::     :::|||                        :::   ::::::
Db         64 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 123

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   |||   :::|||   :::|||      |||:::|||   :::         |||
Db        124 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 183

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                         |||   |||   |||   |||   |||         ||:::
Db        184 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 243

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||     :::   :::        ||||||:::::::::::   :::::::::
Db        244 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 303

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||     :::         .    ||||||:::        ||||||:::
Db        304 GAGGCTCAGTTTGATAGCCGGGTT--------CGTGCAACAGGACACAGTTATGAGAAG 354

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||:::           |||   :::|||   :::   :::         :::|||
Db        355 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 414

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ::::::::::::     ||||||::::::||||||         :::|||:::||||||
Db        415 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 474

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||           |||   ||||||::::::||||||||   ||||||
Db        475 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 531

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173
              ||||||||||||||||||||||||   ||||||:::
Db        532 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 591

Qy        173 ---------------------------------------------------------- 173

Db        592 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 651

Qy        174 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 183
                         |||:::                 :::|||   |||||||||   |||
Db        652 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 711

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::       :::   ||||||||||||||   ||||||||||||
Db        712 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 768

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              |||||||||   |||||||||   :::   |||   ||||||||||||||
Db        769 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 828

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
                 |||||||||   |||   ||||||:::|||   |||:::|||       ::::::

Db        829 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 888

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||:::::::::||||||||||||||||   :::::::::||||||||||||
Db        889 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 948

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                           :::::::::   ||||||:::   :::|||         |||:::
Db        949 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1008

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                 :::     :::         |||:::||||||:::|||   |||:::   |||:::
Db       1009 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1065

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   :::   |||         ||||||   ||||||   ||||||:::||||||||
Db       1066 TATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1125

Qy        324 Thr 324
              |||
Db       1126 ACC 1128

RESULT 13
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
;  GENERAL INFORMATION:
;    APPLICANT:  Hartman, Jacob
;    APPLICANT:  Fulga, Netta
;    APPLICANT:  Mendelovitch, Simona
;    APPLICANT:  Gorecki, Marian
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,760
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/378,233
;      FILING DATE:  25-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
US-08-782-760-5
```

```
Alignment Scores:
Pred. No.:              7.35e-67        Length:         927
Score:                  610.00          Matches:        126
Percent Similarity:     60.50%          Conservative:   44
Best Local Similarity:  44.84%          Mismatches:     71
Query Match:            33.93%          Indels:         41
DB:                     2               Gaps:           5

US-09-980-881A-3 (1-338) x US-08-782-760-5 (1-927)

Qy         93 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 112
              ||||||          |||   :::|||::::::          |||   :::|||||:::
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy        113 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 132
                 :::         :::|||||::::::|||:::      |||||::::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy        133 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 152
                    :::||||||||||||:::   ||||||   :::      |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy        153 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 172
              :::||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy        173 Ile------------------------------------------------------- 173
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297

Qy        174 ----------------------------------GlyHis------------------ 175
                                                  |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy        176 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 195
              :::||||||||||||||   |||||||      |||   :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        196 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 215
                 ||||||||||||   :::   ||||||||   ||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        216 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 235
              ||||||||||||||||||      ||||||||||||   |||   ||||||:::|||
Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        236 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
              |||:::|||   |||::::::   |||||||||||:::::::::|||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        256 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 275
                 :::::::::||||||||||||          :::::::::|||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        276 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 295
                 :::         |||   :::   :::      |||          |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        296 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 315
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        316 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 335
              |||   ||||||||||||||||||||      |||      :::|||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        336 Ser 336
              |||
Db        831 TCT 833

RESULT 14
PCT-US96-00995-5
```

```
; Sequence 5, Application PC/TUS9600995
;  GENERAL INFORMATION:
;    APPLICANT:  Bio-Technology General Corp.
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00995
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847-A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
PCT-US96-00995-5

Alignment Scores:
Pred. No.:              7.35e-67        Length:         927
Score:                  610.00          Matches:        126
Percent Similarity:     60.50%          Conservative:   44
Best Local Similarity:  44.84%          Mismatches:     71
Query Match:            33.93%          Indels:         41
DB:                     5               Gaps:           5

US-09-980-881A-3 (1-338) x PCT-US96-00995-5 (1-927)

Qy         93 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 112
              ||||||          |||   :::|||::::::          |||   :::|||||:::
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy        113 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 132
                 :::         :::|||||::::::|||:::      |||||::::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy        133 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 152
                    :::||||||||||||:::   ||||||   :::      |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy        153 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 172
              :::||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy        173 Ile------------------------------------------------------- 173
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297
```

```
Qy        174 ------------------------------------GlyHis------------------ 175
                                                  |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy        176 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 195
              :::|||||||||||||||   ||||||      |||   :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        196 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 215
                 ||||||||||||   :::      |||||||||   |||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        216 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 235
              ||||||||||||||||||      ||||||||||||   |||   ||||||:::|||
Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        236 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
              |||:::|||   |||::::::   |||||||||||:::::::::|||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 275
                 :::::::::|||||||||||                 :::::::::|||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        276 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 295
                 :::         |||   :::   :::      |||         |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        296 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 315
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        316 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 335
              |||   |||||||||||||||||||||          |||      :::||||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        336 Ser 336
              |||
Db        831 TCT 833

RESULT 15
US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..921
US-08-696-139-3

Alignment Scores:
Pred. No.:                2.42e-63        Length:          921
Score:                    582.00          Matches:         117
Percent Similarity:       59.47%          Conservative:    40
Best Local Similarity:    44.32%          Mismatches:      67
Query Match:              32.37%          Indels:          40
DB:                       1               Gaps:            5

US-09-980-881A-3 (1-338) x US-08-696-139-3 (1-921)

Qy         98 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 117
              ||||||:::|||::::::          |||   :::|||   :::   :::|||
Db         16 TATGAGAAGTACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAA 75

Qy        118 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 137
              :::||||||::::::::::::      ||||||::::::|||           :::|||
Db         76 AATCCAGACCTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATAC 135

Qy        138 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 157
              :::||||||||||   ||||||          |||   ||||||::::::|||||||||
Db        136 CTCCTCAAGGTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGT 192

Qy        158 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 173
                 ||||||||||||||||||||   |||||||||   ||||||:::
Db        193 TTCCATGCCAGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTT 252

Qy        173 ------------------------------------------------------------ 173

Db        253 CTCACCTATGGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTC 312

Qy        174 --------------------GlyHis------------------AsnArgMetTrpArg 180
                                  |||:::                  ||||||||||||||
Db        313 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 372

Qy        181 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 200
              |||   ||||||      |||          ||||||||||||||   ||||||||||||
Db        373 AAGACCCGCTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT 432

Qy        201 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 220
                    ||||||      |||||||||:::      |||   ||||||||||||
Db        433 ---GATGCTGGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGT 489

Qy        221 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240
              |||         |||||||||   |||   ||||||:::|||   |||:::|||   |||
Db        490 GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAAC 549

Qy        241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 260
              ::::::   ||||||||||||:::::::::|||||||||||||||   |||::::::|||
Db        550 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCT 609

Qy        261 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 280
              ||||||||                 :::::::::   ||||||:::   :::|||
Db        610 TATTCCTATGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCT 669
```

```
Qy        281 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 300
              ||||||:::   :::      |||          |||:::||||||:::|||   |||:::
Db        670 GCCGTGAAAGAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCT 726

Qy        301 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 320
                 |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db        727 ACAACAATCTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 786

Qy        321 TyrSerPheThr 324
              ||||||||||||
Db        787 TATTCCTTCACC 798
```

Search completed: January 2, 2004, 09:25:54
Job time : 177.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 20:10:07 ; Search time 822.067 Seconds
(without alignments)
1423.134 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135919_26949/app_query.fasta_1.1628
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980881_@CGN_1_1_704_@runat_31122003_135919_26949
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1732	96.3	1728	10	US-09-880-107-2396	Sequence 2396, Ap
2	1628	90.5	1272	13	US-10-379-836-1	Sequence 1, Appli
3	902	50.2	1400	9	US-09-925-302-24	Sequence 24, Appl
4	622	34.6	1254	15	US-10-229-546-3	Sequence 3, Appli
5	622	34.6	1622	12	US-10-429-802-21	Sequence 21, Appl
6	622	34.6	1622	15	US-10-229-546-1	Sequence 1, Appli
7	622	34.6	1633	13	US-10-341-434-187	Sequence 187, App
8	619.5	34.5	1311	14	US-10-200-344-9	Sequence 9, Appli
9	619.5	34.5	1993	12	US-10-274-639-33	Sequence 33, Appl
10	619.5	34.5	2154	10	US-09-910-059-124	Sequence 124, App
11	618	34.4	1332	10	US-09-954-456-1141	Sequence 1141, Ap
12	591	32.9	416	10	US-09-960-352-14595	Sequence 14595, A
13	571.5	31.8	2128	14	US-10-200-344-13	Sequence 13, Appl
14	569	31.6	1125	9	US-09-888-615-2	Sequence 2, Appli
15	569	31.6	1332	16	US-10-176-306-75	Sequence 75, Appl
16	569	31.6	1603	16	US-10-176-306-73	Sequence 73, Appl
17	562	31.3	1826	13	US-10-252-157-453	Sequence 453, App
18	554	30.8	1200	15	US-10-200-910-7	Sequence 7, Appli
19	549	30.5	1870	10	US-09-910-059-112	Sequence 112, App
20	534.5	29.7	1050	14	US-10-200-344-11	Sequence 11, Appl
21	527.5	29.3	1311	15	US-10-200-910-5	Sequence 5, Appli
22	519.5	28.9	1311	14	US-10-076-535-3	Sequence 3, Appli
23	519.5	28.9	1687	14	US-10-076-535-1	Sequence 1, Appli
24	511.5	28.4	1306	9	US-09-923-779-144	Sequence 144, App
25	511.5	28.4	1306	9	US-09-962-436-557	Sequence 557, App
26	507.5	28.2	1341	9	US-09-925-297-69	Sequence 69, Appl
27	493.5	27.4	1254	9	US-09-925-297-58	Sequence 58, Appl
28	493.5	27.4	1260	13	US-10-345-680-66	Sequence 66, Appl
29	493.5	27.4	1380	10	US-09-969-347-209	Sequence 209, App
30	493.5	27.4	1380	13	US-10-345-680-64	Sequence 64, Appl
31	491	27.3	2786	11	US-09-946-374-233	Sequence 233, App
32	491	27.3	2786	13	US-10-015-387A-233	Sequence 233, App
33	491	27.3	2786	13	US-10-006-130A-233	Sequence 233, App
34	491	27.3	2786	13	US-10-199-672-307	Sequence 307, App
35	491	27.3	2786	13	US-10-006-172A-233	Sequence 233, App
36	491	27.3	2786	13	US-10-187-749-307	Sequence 307, App
37	491	27.3	2786	13	US-10-194-457-307	Sequence 307, App
38	491	27.3	2786	13	US-10-184-642-307	Sequence 307, App
39	491	27.3	2786	13	US-10-196-747-307	Sequence 307, App
40	491	27.3	2786	13	US-10-015-392A-233	Sequence 233, App
41	491	27.3	2786	13	US-10-017-253A-233	Sequence 233, App
42	491	27.3	2786	13	US-10-173-689-307	Sequence 307, App
43	491	27.3	2786	13	US-10-173-690-307	Sequence 307, App
44	491	27.3	2786	13	US-10-173-691-307	Sequence 307, App
45	491	27.3	2786	13	US-10-173-692-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```
;  OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Alignment Scores:
Pred. No.:              5.72e-211       Length:         1728
Score:                  1732.00         Matches:        338
Percent Similarity:     86.01%          Conservative:   0
Best Local Similarity:  86.01%          Mismatches:     0
Query Match:            96.33%          Indels:         55
DB:                     10              Gaps:           2

US-09-980-881A-3 (1-338) x US-09-880-107-2396 (1-1728)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         86 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 505

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        506 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 565

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        566 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 625

Qy        175 ------------------------------------------------------------ 175

Db        626 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 685

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 |||||||||||||||||||||||
Db        686 GTTAATGTGGACGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 745

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        746 TCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 805

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        806 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 865

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        866 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAATATCAACCAG 925

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        926 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTTTCCATATTCCTAT 985

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        986 ACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1045

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1046 GCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGGCTCAGAAACCTTA 1105

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1106 TACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1165

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1166 ACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGAGCGTTACATCAAA 1225

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||||||||||||||||||||||||||||||||||
Db       1226 CCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 1262

RESULT 2
US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1272
;   TYPE: DNA
;   ORGANISM: Papio hamadryas
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1269)
US-10-379-836-1

Alignment Scores:
Pred. No.:              7.23e-198       Length:         1272
Score:                  1628.00         Matches:        318
Percent Similarity:     83.21%          Conservative:   9
Best Local Similarity:  80.92%          Mismatches:     11
Query Match:            90.55%          Indels:         55
DB:                     13              Gaps:           2

US-09-980-881A-3 (1-338) x US-10-379-836-1 (1-1272)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         67 TTTCAGAGTGGCCAGGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTGCTA 126

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        127 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCGGACCTTATT 186

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                 ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db        187 GAGAAGAAAAAACAAGTCCATTTTTTTGTAAATTCATCTGATGTCGACAATGTGAAAGCC 246

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        247 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCCTGCTGGCAGATGTGGAAGATCTTATT 306

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        307 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 366
```

```
Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||   |||||||||:::::|||||
Db        367 TATCACTCACTAAATGAAATCTATTCTTGGATAGAACTTATAACTGAGAAGTATCCTGAT 426

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||:::|||||:::||||||||||||||||||||
Db        427 ATGCTTACAAAAATCCACATTGGATCCTCCTATGAGAAGCACCCACTTTATGTTTTAAAG 486

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db        487 GTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATGTGGATTGACTGTGGAATCCATGCC 546

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        547 AGAGAATGGATCTCCCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTGAATACTAC 606

Qy        175 ------------------------------------------------------------ 175

Db        607 GGGATAATAGGGGAATATACCAATCTTCTGAGGCATGTGGATTTCTATGTTATGCCAGTG 666

Qy        176 -----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  ||||||||||||||||||||||
Db        667 GTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 726

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        727 TCTTTCTATGCGAACAATCGTTGCATCGGAACAGACCTGAACAGGAACTTTGCGTCCAAA 786

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db        787 CACTGGTGTGAGGAAGGTGCATCCAGTTTCTCATGCTCGGAAACCTACTGTGGACTTTAT 846

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              ||||||||||||||||||   ||||||||||:::||||||||||||||||||||||||||
Db        847 CCTGAGTCAGAACCAGAAGCGAAGGCGGTGGCTAATTTCTTGAGAAGAAATATCAACCAC 906

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSerTyr 263
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 ATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATCGTGTTTCCATATTCCTAT 966

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 ACTCGAAGCAAAAGCAAAGACCACGAGGAATTGTCTCTAGTAGCCAGTGAAGCAGTTCGT 1026

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
              ||||||:::|||||||||||||||   |||||||||||||   |||||||||||||||||
Db       1027 GCTATTCAGAAAACCAGTAAAAATATCAGGTATACACATGGCCGTGGCTCAGAAACCTTA 1086

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db       1087 TACCTAGCTCCTGGAGGTGCGGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1146

Qy        324 ------------------------------------------------------ThrSerAs 326
                                                                     ||||||||
Db       1147 ACAATTGAACTTCGAGATACGGGCAAATACGGATTCTTGCTGCCTGAGCGTTACATCAAA 1206

Qy        326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
              ||||   |||:::   ||||||||||||||||||||
Db       1207 CCCACTTGTAAAGACGCTTTTGCCGCTGTCTCTAAAA 1243

RESULT 3
US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
;   LENGTH: 1400
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (25)
;   OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Alignment Scores:
Pred. No.:              8.81e-105       Length:         1400
Score:                  902.00          Matches:        181
Percent Similarity:     76.05%          Conservative:   0
Best Local Similarity:  76.05%          Mismatches:     1
Query Match:            50.17%          Indels:         56
DB:                     9               Gaps:           2

US-09-980-881A-3 (1-338) x US-09-925-302-24 (1-1400)

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 175
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db         10 GGAATCCATGCCAGANAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATA 69

Qy        175 ------------------------------------------------------------ 175

Db         70 ACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTAT 129

Qy        176 ------------------------------------------------AsnArgMetTr 179
                                                             ||||||||||
Db        130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACTACTCATGGAAAAAGAATCGAATGTG 189

Qy        179 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 199
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAA 249

Qy        199 nPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 219
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309

Qy        219 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 239
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAG 369

Qy        239 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 259
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 AAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTT 429

Qy        259 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 279
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 TCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAG 489

Qy        279 rGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 299
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 TGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGG 549

Qy        299 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIl 319
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609

Qy        319 eLysTyrSerPhe---------------------------------------------- 323
              ||||||||||||
Db        610 CAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGA 669

Qy        324 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 338
                   ||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 719

RESULT 4
```

```
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1254
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1254)
US-10-229-546-3

Alignment Scores:
Pred. No.:              5.48e-69        Length:         1254
Score:                  622.00          Matches:        128
Percent Similarity:     55.12%          Conservative:   71
Best Local Similarity:  35.46%          Mismatches:     118
Query Match:            34.59%          Indels:         44
DB:                     15              Gaps:           7

US-09-980-881A-3 (1-338) x US-10-229-546-3 (1-1254)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||         :::|||           |||:::     :::|||      ::::::
Db         61 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 120

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::::|||      |||   |||:::     |||   |||                 :::
Db        121 AAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTA 180

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                       |||   |||   |||:::     :::          :::::::::
Db        181 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 240

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
                 |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db        241 GCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 300

Qy         81 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 99
              :::::|||           :::   :::      |||   |||          |||
Db        301 GAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCA 351

Qy        100 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 119
              :::|||::::::     :::|||   :::|||   |||   :::   :::::::::|||
Db        352 AAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCT 411

Qy        120 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 139
              :::|||:::::::::|||   |||||||||:::   |||      ||||||||||||||
Db        412 GAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTG 471

Qy        140 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 159
              |||:::   |||::::::          :::   ||||||::::::||||||||||||||
Db        472 AAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCAC 528

Qy        160 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle----------------- 173
              ||||||||||||:::|||||||||||||||   ||||||:::
Db        529 GCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACT 588

Qy        173 ------------------------------------------------------------ 173
```

```
Db        589 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 648

Qy        174 ---------------GlyHis-----------------AsnArgMetTrpArgLysAsn 182
                             |||:::                |||||||||||||||||||
Db        649 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 708

Qy        183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
              ||||||          |||:::   |||||||||||||||||||||||||||||   :::
Db        709 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 765

Qy        203 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 222
                    |||          ::::::      |||::::::   |||   |||
Db        766 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825

Qy        223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
                |||||||||||||   |||   |||||||||   :::|||:::|||   :::::::|||
Db        826 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 885

Qy        243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 262
              :::||||||   ||||||:::   |||||||||||||||   ::::::|||||||||
Db        886 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 945

Qy        263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
              ||||||                ::::||||||:::|||:::   ||||||
Db        946 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1005

Qy        283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
                    :::      |||         |||||||||   :::|||            |||
Db       1006 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1062

Qy        303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
              :::|||          |||         ||||||   |||||||||||||||||::::::
Db       1063 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1122

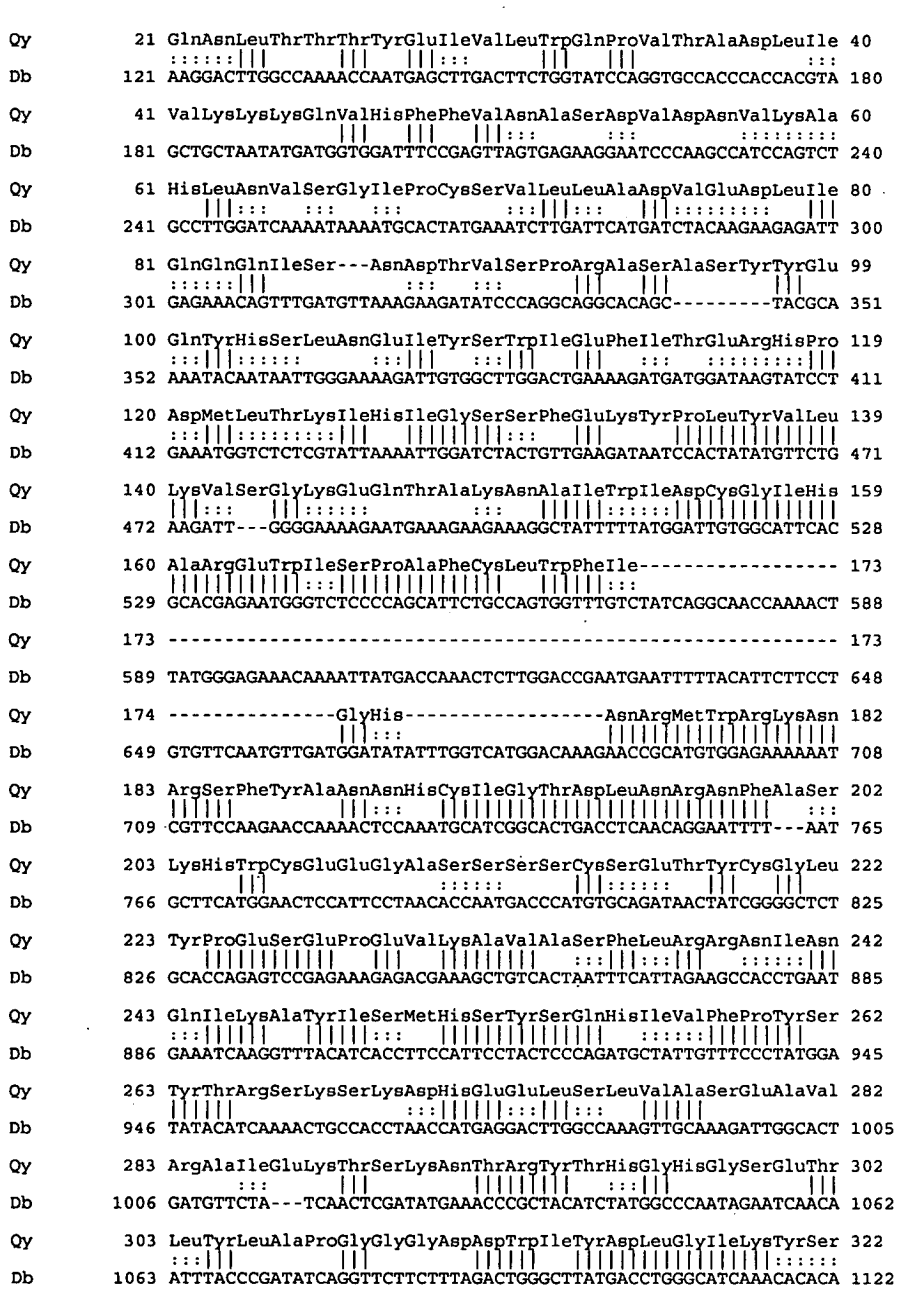
Qy        323 Phe 323
              |||
Db       1123 TTT 1125
```

```
RESULT 5
US-10-429-802-21
; Sequence 21, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
;  APPLICANT: HUNG, MIEN-CHIE
;  APPLICANT:  WONG, KA YIN
;  APPLICANT:  ZOU, YIYU
;  TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
;  FILE REFERENCE: UTSC:752US
;  CURRENT APPLICATION NUMBER: US/10/429,802
;  CURRENT FILING DATE:  2003-05-05
;  PRIOR APPLICATION NUMBER: 60/377,672
;  PRIOR FILING DATE: 2002-05-03
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 1622
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-429-802-21

Alignment Scores:
Pred. No.:              8.43e-69        Length:         1622
Score:                  622.00          Matches:        128
Percent Similarity:     55.12%          Conservative:   71
Best Local Similarity:  35.46%          Mismatches:     118
Query Match:            34.59%          Indels:         44
DB:                     12              Gaps:           7

US-09-980-881A-3 (1-338) x US-10-429-802-21 (1-1622)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||         :::|||           |||:::     :::|||      ::::::
Db         61 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 120
```

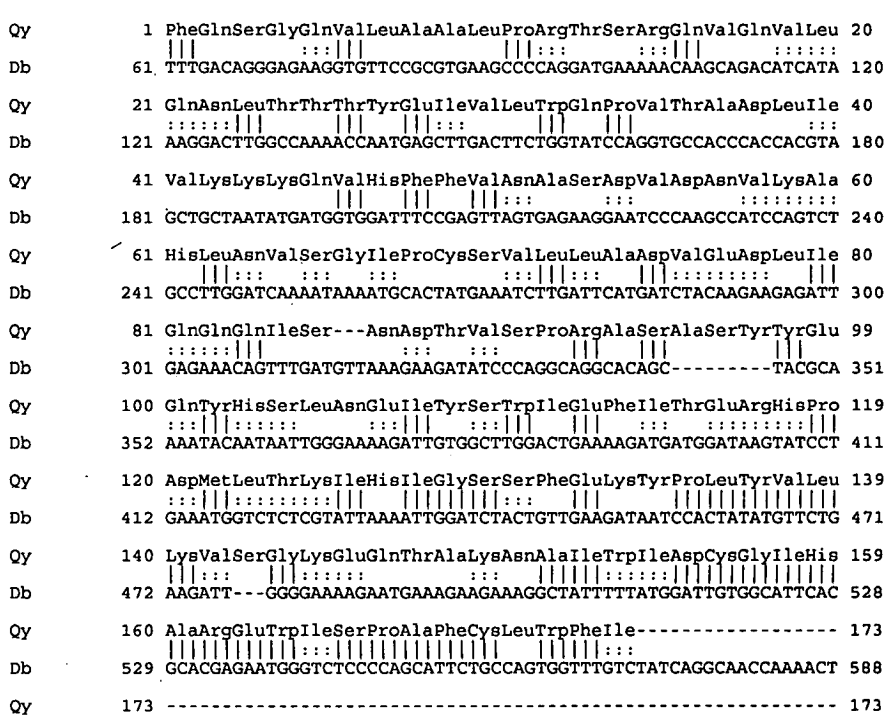


RESULT 6
US-10-229-546-1
; Sequence 1, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MPI01-156P1RNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1622
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-1

Alignment Scores:

Pred. No.:	8.43e-69	Length:	1622
Score:	622.00	Matches:	128
Percent Similarity:	55.12%	Conservative:	71
Best Local Similarity:	35.46%	Mismatches:	118
Query Match:	34.59%	Indels:	44
DB:	15	Gaps:	7

US-09-980-881A-3 (1-338) x US-10-229-546-1 (1-1622)



```
Db        589 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 648

Qy        174 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 182
                          |||:::                 ||||||||||||||||||||
Db        649 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 708

Qy        183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
              ||||||         |||:::  ||||||||||||||||||||||||||||||   :::
Db        709 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 765

Qy        203 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 222
                  |||                  ::::::      |||::::::   |||   |||
Db        766 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 825

Qy        223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
                 ||||||||||||   |||   |||||||||   :::|||:::|||   :::::|||
Db        826 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 885

Qy        243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 262
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        886 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 945

Qy        263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
              ||||||                ::::||||||:::|||:::   ||||||
Db        946 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1005

Qy        283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
                  :::      |||          |||||||||   :::|||                |||
Db       1006 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1062

Qy        303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
              :::|||          |||          ||||||   ||||||||||||||||||::::::
Db       1063 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1122

Qy        323 Phe 323
              |||
Db       1123 TTT 1125
```

RESULT 7
US-10-341-434-187
; Sequence 187, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
; LENGTH: 1633
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)..(1262)
; OTHER INFORMATION:
US-10-341-434-187

Alignment Scores:

Pred. No.:	8.53e-69	Length:	1633
Score:	622.00	Matches:	128
Percent Similarity:	55.12%	Conservative:	71
Best Local Similarity:	35.46%	Mismatches:	118
Query Match:	34.59%	Indels:	44
DB:	13	Gaps:	7

US-09-980-881A-3 (1-338) x US-10-341-434-187 (1-1633)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||        :::|||           |||:::     :::|||      ::::::
Db         72 TTTGACAGGGAGAAGGTGTTCCGCGTGAAGCCCCAGGATGAAAAACAAGCAGACATCATA 131

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ::::::|||      |||   |||:::     |||   |||                :::
Db        132 AAGGACTTGGCCAAAACCAATGAGCTTGACTTCTGGTATCCAGGTGCCACCCACCACGTA 191

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                    |||   |||   |||:::      :::        :::::::::
Db        192 GCTGCTAATATGATGGTGGATTTCCGAGTTAGTGAGAAGGAATCCCAAGCCATCCAGTCT 251

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||:::   :::   :::         :::|||:::   |||:::::::::   |||
Db        252 GCCTTGGATCAAAATAAAATGCACTATGAAATCTTGATTCATGATCTACAAGAAGAGATT 311

Qy         81 GlnGlnGlnIleSer---AsnAspThrValSerProArgAlaSerAlaSerTyrTyrGlu 99
              ::::::|||            :::   :::      |||   |||         |||
Db        312 GAGAAACAGTTTGATGTTAAAGAAGATATCCCAGGCAGGCACAGC---------TACGCA 362

Qy        100 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 119
              :::|||::::::     :::|||  :::|||   |||   :::   :::::::::|||
Db        363 AAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAGTATCCT 422

Qy        120 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 139
              :::|||:::::::::|||   |||||||||:::   |||       ||||||||||||||
Db        423 GAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATATGTTCTG 482

Qy        140 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 159
              |||:::   |||::::::          :::   ||||||:::::|||||||||||||||
Db        483 AAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGCATTCAC 539

Qy        160 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 173
              ||||||||||:::|||||||||||||||   ||||||:::
Db        540 GCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACCAAAACT 599

Qy        173 ------------------------------------------------------------ 173

Db        600 TATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATTCTTCCT 659

Qy        174 --------------GlyHis-----------------AsnArgMetTrpArgLysAsn 182
                          |||:::                 ||||||||||||||||||||
Db        660 GTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGAAAAAAT 719

Qy        183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
              ||||||         |||:::  ||||||||||||||||||||||||||||||   :::
Db        720 CGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT---AAT 776

Qy        203 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 222
                  |||                  ::::::      |||::::::   |||   |||
Db        777 GCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGGGGCTCT 836

Qy        223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
                 ||||||||||||   |||   |||||||||   :::|||:::|||   :::::|||
Db        837 GCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCACCTGAAT 896

Qy        243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSer 262
              :::||||||   ||||||:::   ||||||||||||||   ::::::|||||||||
Db        897 GAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCCTATGGA 956

Qy        263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
              ||||||                ::::||||||:::|||:::   ||||||
Db        957 TATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATTGGCACT 1016

Qy        283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
                  :::      |||          |||||||||   :::|||                |||
Db       1017 GATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAATCAACA 1073

Qy        303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
              :::|||          |||          ||||||   ||||||||||||||||||::::::
Db       1074 ATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAACACACA 1133
```

```
Qy        323 Phe 323
              |||
Db       1134 TTT 1136

RESULT 8
US-10-200-344-9
; Sequence 9, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 1311
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-200-344-9

Alignment Scores:
Pred. No.:                1.23e-68      Length:          1311
Score:                    619.50        Matches:         136
Percent Similarity:       51.15%        Conservative:    64
Best Local Similarity:    34.78%        Mismatches:      127
Query Match:              34.45%        Indels:          64
DB:                       14            Gaps:            8

US-09-980-881A-3 (1-338) x US-10-200-344-9 (1-1311)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              :::         :::|||:::      :::|||:::|||      :::          |||
Db        109 TATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTG 168

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   ::::::        ::::::   ||||||||||||   :::          :::
Db        169 AAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCCAGCAGTATCTCCTATGTA 228

Qy         41 ValLys------LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 58
                 :::               ||||||         |||   |||              :::
Db        229 TCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCCGAGCC------CTG 282

Qy         59 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 78
                 |||   |||       :::   |||          ||||||:::   |||::::::
Db        283 TTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAGAAGATCTTCAGAAA 342

Qy         79 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer------ 96
                 :::::::::       |||       |||          |||   |||   |||
Db        343 ACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGATCCCTCTCTGGATAT 402

Qy         97 TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 116
                 ||||||   ||||||||||||   ||||||   :::|||:::      :::   :::
Db        403 AATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAA 462

Qy        117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 136
                 |||       ::::::          ||||||   |||:::|||          |||
Db        463 ACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTT 522

Qy        137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCys 156
                 ::::::||||||:::   |||:::          |||   |||:::||||||||||||
Db        523 TTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGT 579

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle--------- 173
              |||||||||||||||||||||||    ||||||||||||   ||||||:::
Db        580 GGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCT 639

Qy        173 ------------------------------------------------------------ 173

Db        640 CTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTAT 699

Qy        174 ----------------------------------------GlyHisAsnArgMetTrp 179
                                                      ::::::|||   |||
Db        700 ATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGG 759

Qy        180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199
              ||||||   ||||||       :::          |||   |||   |||   |||||||||
Db        760 AGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAAC 819

Qy        200 PheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyr 219
              :::       |||   ||||||:::||||||||||||          |||   :::||||||
Db        820 TGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATAC 876

Qy        220 CysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArg 239
              ||||||   :::||||||||||||||||||||||||||||||||:::|||||||||:::
Db        877 TGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAA 936

Qy        240 AsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhe 259
              :::         |||:::||||||:::|||   |||:::|||:::|||   :::::::::
Db        937 CACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTAT 996

Qy        260 ProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSer 279
              ||||||||||||       :::          :::          :::          |||
Db        997 CCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTAT 1056

Qy        280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 299
              :::||||||   |||::::::                ||||||   :::|||
Db       1057 AAAGCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCC 1113

Qy        300 SerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIle 319
              |||   |||||||||::::::   |||          ||||||   |||      ||||||
Db       1114 TCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATA 1173

Qy        320 LysTyrSerPhe----------------------------------------------- 323
                 |||:::|||
Db       1174 CCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAG 1233

Qy        324 ----ThrSerAsnProProValGluLysLeu 332
                  :::||||||||||||||:::||||||
Db       1234 ATGCTCATCAAACCCACCTGTACAGAAACTA 1264

RESULT 9
US-10-274-639-33
; Sequence 33, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
```

```
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 33
;   LENGTH: 1993
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CB1
US-10-274-639-33

Alignment Scores:
Pred. No.:              2.49e-68        Length:         1993
Score:                  619.50          Matches:        136
Percent Similarity:     51.15%          Conservative:   64
Best Local Similarity:  34.78%          Mismatches:     127
Query Match:            34.45%          Indels:         64
DB:                     12              Gaps:           8

US-09-980-881A-3 (1-338) x US-10-274-639-33 (1-1993)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              :::     :::|||:::     :::|||:::|||     :::         |||
Db        446 TATGCTGGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTG 505

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   ::::::         ::::::   ||||||||||||   :::         :::
Db        506 AAGAAAATATCCTATCAACTTAAGGTGGACCTGTGGCAGCCCAGCAGTATCTCCTATGTA 565

Qy         41 ValLys------LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnVal 58
                 :::              ||||||          |||   |||             :::
Db        566 TCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCCGAGCC------CTG 619

Qy         59 LysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAsp 78
              |||   |||      :::   |||        ||||||:::   |||::::::
Db        620 TTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAGAAGATCTTCAGAAA 679

Qy         79 LeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSer------ 96
                 :::::::::      |||      |||          |||   |||   |||
Db        680 ACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGATCCCTCTCTGGATAT 739

Qy         97 TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGlu 116
              ||||||   ||||||||||||   ||||||   :::|||:::      :::   :::
Db        740 AATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAA 799

Qy        117 ArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeu 136
              |||      ::::::          ||||||   |||:::|||          |||
Db        800 ACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTT 859

Qy        137 TyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCys 156
              ::::::||||||:::   |||:::             |||   |||:::|||||||||||||
Db        860 TTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGT 916

Qy        157 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle--------- 173
              ||||||||||||||||||||   ||||||||||||   ||||||:::
Db        917 GGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCT 976

Qy        173 ------------------------------------------------------------ 173

Db        977 CTTCTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTAT 1036

Qy        174 ----------------------------------------GlyHisAsnArgMetTrp 179
                                                     ::::::|||   |||
Db       1037 ATCATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGG 1096

Qy        180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsn 199
              ||||||   ||||||      :::         |||   |||   |||   |||||||||
Db       1097 AGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAAC 1156

Qy        200 PheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyr 219
              :::      |||   ||||||:::||||||||||||          |||   :::||||||
Db       1157 TGGAAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATAC 1213

Qy        220 CysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArg 239
              ||||||   :::||||||||||||||||||||||||||||||:::|||||||||:::
Db       1214 TGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAA 1273

Qy        240 AsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhe 259
              :::         |||:::||||||:::|||   |||:::|||:::|||   :::::::::
Db       1274 CACAGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTAT 1333

Qy        260 ProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSer 279
              ||||||||||      :::         :::         :::         |||
Db       1334 CCCTATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTAT 1393

Qy        280 GluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGly 299
              :::||||||   |||::::::                ||||||   :::|||
Db       1394 AAAGCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCC 1450

Qy        300 SerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIle 319
              |||   |||||||||:::::   |||         ||||||   |||      ||||||
Db       1451 TCCACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATA 1510

Qy        320 LysTyrSerPhe----------------------------------------------- 323
                 |||:::|||
Db       1511 CCTTATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAG 1570

Qy        324 ----ThrSerAsnProProValGluLysLeu 332
                  :::|||||||||||||||:::||||||
Db       1571 ATGCTCATCAAACCCACCTGTACAGAAACTA 1601

RESULT 10
US-09-910-059-124
; Sequence 124, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody, a
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 124
;   LENGTH: 2154
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
```

US-09-910-059-124

Alignment Scores:

Pred. No.:	2.83e-68	Length:	2154
Score:	619.50	Matches:	127
Percent Similarity:	54.29%	Conservative:	69
Best Local Similarity:	35.18%	Mismatches:	122
Query Match:	34.45%	Indels:	43
DB:	10	Gaps:	6

US-09-980-881A-3 (1-338) x US-09-910-059-124 (1-2154)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::     :::|||                              :::   ::::::
Db         64 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 123

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   |||   :::|||   :::|||      |||:::|||   :::        |||
Db        124 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 183

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                        |||   |||   |||   |||   |||          |||:::
Db        184 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 243

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||      :::  :::         ||||||:::::::::::   :::::::::
Db        244 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 303

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::   |||      :::              ||||||:::         ||||||:::
Db        304 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 354

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              |||:::             |||   :::|||   :::   :::         :::|||
Db        355 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 414

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ::::::::::::      ||||||::::::||||||        :::|||:::||||||
Db        415 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 474

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||   ||||||           |||   ||||||::::::||||||||   ||||||
Db        475 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 531

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173
              ||||||||||||||||||||||||   ||||||:::
Db        532 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 591

Qy        173 ------------------------------------------------------------ 173

Db        592 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 651

Qy        174 ------------GlyHis-----------------AsnArgMetTrpArgLysAsnArg 183
                         |||:::                 :::|||   |||||||||   |||
Db        652 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 711

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::      :::   ||||||||||||||   ||||||||||||
Db        712 TCCACCCATACTGGATCTAGCTGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 768

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              |||||||||   |||||||||   :::   |||   ||||||||||||||||
Db        769 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 828

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
              |||||||||   |||   ||||||:::|||   |||:::|||      ::::::
Db        829 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 888

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 263
              ||||||||||:::::::::||||||||||||||   :::::::::||||||||||||
Db        889 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 948

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
```

```
                          :::::::::   ||||||:::   :::|||         |||:::
Db        949 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1008

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
                 :::      :::         |||:::||||||:::|||   |||:::   |||:::
Db       1009 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1065

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
              |||   :::   |||         ||||||   ||||||   ||||||:::|||||||||
Db       1066 TATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1125

Qy        324 Thr 324
              |||
Db       1126 ACC 1128
```

RESULT 11
US-09-954-456-1141
; Sequence 1141, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cance
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1141
; LENGTH: 1332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1141

Alignment Scores:

Pred. No.:	1.97e-68	Length:	1332
Score:	618.00	Matches:	129
Percent Similarity:	54.57%	Conservative:	68
Best Local Similarity:	35.73%	Mismatches:	120
Query Match:	34.37%	Indels:	44
DB:	10	Gaps:	7

US-09-980-881A-3 (1-338) x US-09-954-456-1141 (1-1332)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||:::     :::|||                                 :::   ::::::
Db         87 TTTGAAGGCGAGAAGGTGTTCCGTGTTAACGTTGAAGATGAAAATCACATTAACATAATC 146

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              :::   |||   :::|||   :::|||      |||:::|||   :::        |||
Db        147 CGCGAGTTGGCCAGCACGACCCAGATTGACTTCTGGAAGCCAGATTCTGTCACACAAATC 206

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
```

```
Db        207 AAACCTCACAGTACAGTTGACTTCCGTGTTAAAGCAGAAGATACTGTCACTGTGGAGAAT 266

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Db        267 GTTCTAAAGCAGAATGAACTACAATACAAGGTACTGATAAGCAACCTGAGAAATGTGGTG 326

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Db        327 GAGGCTCAGTTTGATAGCCGGGTT---------CGTGCAACAGGACACAGTTATGAGAAG 377

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Db        378 TACAACAAGTGGGAAACGATAGAGGCTTGGACTCAACAAGTCGCCACTGAGAATCCAGCC 437

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Db        438 CTCATCTCTCGCAGTGTTATCGGAACCACATTTGAGGGACGCGCTATTTACCTCCTGAAG 497

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Db        498 GTT---GGCAAAGCTGGACAAAATAAGCCTGCCATTTTCATGGACTGTGGTTTCCATGCC 554

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------------------- 173

Db        555 AGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTTGTAAGAGAGGCTGTTCGTACCTAT 614

Qy        173 ---------------------------------------------------------- 173

Db        615 GGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAGTTAGACTTTTATGTCCTGCCTGTG 674

Qy        174 ------------GlyHis------------------AsnArgMetTrpArgLysAsnArg 183

Db        675 CTCAATATTGATGGCTACATCTACACCTGGACCAAGAGCCGATTTTGGAGAAAGACTCGC 734

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203

Db        735 TCCACCCATACTGGATCT---AGCATTGGCACAGACCCCAACAGAAATTTT---GATGCT 788

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223

Db        789 GGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGTGATGAAACTTACTGTGGACCTGCC 848

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243

Db        849 GCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGATTTCATCCGCAACAAACTCTCTTCC 908

Qy        244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 263

Db        909 ATCAAGGCATATCTGACAATCCACTCGTACTCCCAAATGATGATCTACCCTTACTCATAT 968

Qy        264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283

Db        969 GCTTACAAACTCGGTGAGAACAATGCTGAGTTGAATGCCCTGGCTAAAGCTACTGTGAAA 1028

Qy        284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303

Db       1029 GAACTT---GCCTCACTGCACGGCACCAAGTACACATATGGCCCGGGAGCTACAACAATC 1085

Qy        304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323

Db       1086 TATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTATGACCAAGGAATCAGATATTCCTTC 1145

Qy        324 Thr 324

Db       1146 ACC 1148

RESULT 12
US-09-960-352-14595
; Sequence 14595, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT: Tao, Nengbing
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 16511.006/37-21(10298)C
;  CURRENT APPLICATION NUMBER: US/09/960,352
;  CURRENT FILING DATE:  2001-09-24
;  NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14595
;   LENGTH: 416
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 62-LIB34-032-Q1-E1-H10
US-09-960-352-14595

Alignment Scores:
Pred. No.:                8.12e-66      Length:          416
Score:                    591.00        Matches:         111
Percent Similarity:       89.13%        Conservative:    12
Best Local Similarity:    80.43%        Mismatches:      15
Query Match:              32.87%        Indels:          0
DB:                       10            Gaps:            0

US-09-980-881A-3 (1-338) x US-09-960-352-14595 (1-416)

Qy         29 IleValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPhe 48

Db          2 ATTGTTCTCTGGCAGCCAGTAGCAGCTGAATATATTGTGAAGGGATACGAAGTCCATTTT 61

Qy         49 PheValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIlePro 68

Db         62 TTTGTGAATGCATCTGATGTAAGCAATGTGAAAGCCCATTTAAATGCGAGCAGAATTCCA 121

Qy         69 CysSerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThr 88

Db        122 TTCAGGGTCCTGGTGGAAAATGTGGAAGATCTTATCCGGCAGCAGACTTCCAATGACACC 181

Qy         89 ValSerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyr 108

Db        182 ATCAGCCCCCGGGCATCCTCCTCCTACTATGAACAGTATCACTCACTAAATGAGATCTAT 241

Qy        109 SerTrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGly 128

Db        242 TCTTGGATAGAAGTTATGACTGAGCGGTATCCTGATATGGTTGAAAAAATCCACATTGGA 301

Qy        129 SerSerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAla 148

Db        302 TCCTCATACGAGAAGTACCCACTTTATGTTTTAAAGGTTTCTAAAAAAGAACAAAGGGCC 361

Qy        149 LysAsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerPro 166

Db        362 AAAAATGCCATGTGGATTGACTGTGGAATCCACGCCAGAGAGTGGATCTCTCCT 415

RESULT 13
US-10-200-344-13
; Sequence 13, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
```

```
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 2128
;   TYPE: DNA
;   ORGANISM: homo sapiens
US-10-200-344-13

Alignment Scores:
Pred. No.:                3.92e-62        Length:          2128
Score:                    571.50          Matches:         140
Percent Similarity:       45.35%          Conservative:    60
Best Local Similarity:    31.75%          Mismatches:      125
Query Match:              31.79%          Indels:          116
DB:                       14              Gaps:            10

US-09-980-881A-3 (1-338) x US-10-200-344-13 (1-2128)

Qy          3 SerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeuGlnAsn 22
              |||   :::|||:::      :::|||:::|||      :::         |||:::
Db        461 AGTGATAAAGTGATAAGATTTATTCCCAAAACAGAAGAGGAAGCATATGCACTGAAGAAA 520

Qy         23 LeuThr-ThrThrTyrGluIleVal-----------------LeuTrpGlnProValTh 36
              ::::::             :::|||                 ||||||||||||   ::
Db        521 ATATCCTATCAACTTAAGGTTGGTTCCTGCACTACAGGTGGACCTGTGGCAGCCCAGCAG 580

Qy         36 rAlaAspLeuIleValLys------LysLysGlnValHisPheValAsnAlaSerAs 54
              :      :::  :::                |||||      |||   |||
Db        581 TATCTCCTATGTATCAGAGGGAACAGTTACTGATGTCCATATCCCCCAAAATGGTTCCCG 640

Qy         54 pValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCysSerValLeuLeuAl 74
                  :::   |||   |||       :::   |||          ||||||:::
Db        641 AGCC------CTGTTAGCCTTCTTACAGGAAGCCAACATCCAGTACAAGGTCCTCATAGA 694

Qy         74 aAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSe 94
               |||::::::     :::::::::      |||      |||        |||   ||
Db        695 AGATCTTCAGAAAACACTGGAGAAGGGAAGCAGCTTGCACACCCAGAGAAACCGAAGATC 754

Qy         94 rAlaSer------TyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGl 112
              |   |||        ||||||  ||||||||||||  ||||||   :::|||:::
Db        755 CCTCTCTGGATATAATTATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCA 814

Qy        112 uPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGl 132
                  :::   :::   |||      ::::::           ||||||   |||:::||
Db        815 TCATCTGAATAAAACTCACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGA 874

Qy        132 uLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIl 152
              |         |||::::::||||||:::   |||:::            |||   |||::
Db        875 GGGAAGATCTCTTTTTATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGT 931

Qy        152 eTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPh 172
              :||||||||||||||||||||||||||||||||   ||||||||||   |||||
Db        932 TTGGATAGACTGTGGTATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTT 991

Qy        172 eIle--------------------GlyHis-Asn------------------------ 176
              |:::                    ||| |||
Db        992 TGTAAAAGAAGTCCTAGAAAACACAGCTCACAAATGTCAAGAATGTACTAAATTTACAAA 1051

Qy        176 ------------------------------------------------------------ 176

Db       1052 ATATCTCTGCCACTACCAAAACCACAAAAGTATGCTTAATCTTGTAAGTATTGAGTAATA 1111

Qy        176 ------------------------------------------------------------ 176

Db       1112 AAATTTTCTAAACATTCCTAAAAACTCTTCTAACATATAAGAGTGACCCAGCCATGAGAA 1171

Qy        176 ------------------------------------------------------------ 176

Db       1172 AAATGTTGAATCATCTATATTTCTATATCATGCCTGTGTTTAACGTCGATGGATACCATT 1231

Qy        177 ----------------ArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisC 191
                           |||   |||||||||  ||||||      :::         |
Db       1232 TTAGTTGGACCAATGATCGATTTTGGAGAAAAACAAGGTCAAGGAACTCAAGGTTTCGCT 1291

Qy        191 ysIleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaS 211
              ||   |||   |||   |||||||||:::      |||   ||||||:::||||||||||
Db       1292 GCCGTGGAGTGGATGCCAATAGAAACTGGAAAGTGAAG---TGGTGTGATGAAGGAGCTT 1348

Qy        211 erSerSerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValL 231
              ||          |||   :::||||||||||||   :::||||||||||||||||||||
Db       1349 CTATGCACCCTTGTGATGACACATACTGTGGCCCTTTTCCAGAATCTGAGCCGGAAGTGA 1408

Qy        231 ysAlaValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetH 251
              ||||||||||:::|||||||||::::::         |||:::||||||:::|||   |
Db       1409 AGGCTGTAGCTAACTTCCTTCGAAAACACAGAAAGCACATTAGGGCTTATCTCTCCTTTC 1468

Qy        251 isSerTyrSerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspH 271
              ||:::|||:::|||   :::::::::||||||||||||      :::         :::
Db       1469 ATGCATATGCTCAGATGTTACTGTATCCCTATTCTTACAAATATGCAACAATTCCCAATT 1528

Qy        271 isGluGluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysA 291
                        :::      |||   :::||||||   |||::::::
Db       1529 TTAGATGTGTGGAATCTGCAGCTTATAAAGCTGTGAATGCACTTCAGTCAGTA---TACG 1585

Qy        291 snThrArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyA 311
                    ||||||   :::|||       |||   |||||||||::::::   |||
Db       1586 GGGTACGATACAGATATGGACCAGCCTCCACAACGTTGTATGTGAGCTCTGGTAGCTCAA 1645

Qy        311 spAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe---------------------- 323
                  ||||||   |||       ||||||   |||:::|||
Db       1646 TGGATTGGGCCTACAAAAATGGAATACCTTATGCATTTGCTTTCGAACTACGTGACACTG 1705

Qy        324 ------------------------------ThrSerAsnProProValGluLysLeu 332
                                           :::||||||||||||||:::||||||
Db       1706 GATATTTTGGATTTTTACTCCCAGAGATGCTCATCAAACCCACCTGTACAGAAACTA 1762

RESULT 14
US-09-888-615-2
; Sequence 2, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1125
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-888-615-2

Alignment Scores:
Pred. No.:                2.82e-62        Length:          1125
Score:                    569.00          Matches:         110
Percent Similarity:       57.09%          Conservative:    51
Best Local Similarity:    39.01%          Mismatches:      81
Query Match:              31.65%          Indels:          40
DB:                       9               Gaps:            5

US-09-980-881A-3 (1-338) x US-09-888-615-2 (1-1125)

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr---TyrGlu 99
              :::|||:::|||   :::   :::||||||||||   :::      :::|||   |||
Db         85 AGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAAC 144

Qy        100 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 119
                  ||||||   :::   |||||||||   |||:::      |||:::|||::::::
```

```
Db        145 ATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAG 204

Qy        120 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 139
              :::::::::|||:::      :::|||   :::::::|||   :::|||:::|||   |||
Db        205 GAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTATTATCTG 264

Qy        140 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 159
              |||:::|||                  |||      |||||||:::|||||||||||||||
Db        265 AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCAC 324

Qy        160 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 173
              |||||||||||||||:::|||||||||||   ||||||:::
Db        325 GCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAAC 384

Qy        173 ------------------------------------------------------------ 173

Db        385 CATAAAGACAACTCAAGTATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCA 444

Qy        174 ---------------GlyHis-----------------AsnArgMetTrpArgLysAsn 182
                              |||:::                 :::|||:::||||||||:::
Db        445 GTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC 504

Qy        183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
              ||||||   :::   |||      |||   ||||||||||||||||||||||   :::
Db        505 CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC---AAT 561

Qy        203 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 222
                  ||||||      |||||||||   :::         :::|||:::||||||
Db        562 GCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACA 621

Qy        223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
                 |||   ||||||||||||   ||||||||||||||||||:::             :::
Db        622 GGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGAT 681

Qy        243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 262
                 |||      :::::::::||||||||||||   |||   |||:::   ||||||
Db        682 GATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGC 741

Qy        263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
              ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db        742 TACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCA 801

Qy        283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
                 |||:::   |||           |||   |||      |||      ::::::
Db        802 AATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATT 858

Qy        303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
              ||||||   :::   |||           ||||||     |||:::||||||   :::|||
Db        859 TTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCA 918

Qy        323 PheThr 324
              :::|||
Db        919 TATACG 924
```

```
RESULT 15
US-10-176-306-75
; Sequence 75, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 75
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-176-306-75

Alignment Scores:
Pred. No.:                3.74e-62        Length:          1332
Score:                    569.00          Matches:         110
Percent Similarity:       57.09%          Conservative:    51
Best Local Similarity:    39.01%          Mismatches:      81
Query Match:              31.65%          Indels:          40
DB:                       16              Gaps:            5

US-09-980-881A-3 (1-338) x US-10-176-306-75 (1-1332)

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyr---TyrGlu 99
              :::|||:::|||   :::   :::|||||||||   :::      :::|||   |||
Db        475 AGACAAGAGATTGTGGACAAGTCAGTGAGTCCATGGAGCCTGGAGACGTATTCCTATAAC 534

Qy        100 GlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisPro 119
                 ||||||   :::   |||||||||   |||:::      |||:::|||::::::
Db        535 ATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAGTACAAG 594

Qy        120 AspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeu 139
              :::::::::|||:::      :::|||   :::::::|||   :::|||:::|||   |||
Db        595 GAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATATATTATCTG 654

Qy        140 LysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHis 159
              |||:::|||                  |||      |||||||:::|||||||||||||||
Db        655 AAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGAATTCAC 714

Qy        160 AlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------------ 173
              |||||||||||||||:::|||||||||||   ||||||:::
Db        715 GCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTACAAAAC 774

Qy        173 ------------------------------------------------------------ 173

Db        775 CATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTCCTTCCA 834

Qy        174 ---------------GlyHis-----------------AsnArgMetTrpArgLysAsn 182
                              |||:::                 :::|||:::||||||||:::
```

```
Db        835 GTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGGAAATCC 894

Qy        183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 202
              ||||||   :::   |||       |||   ||||||||||||||||||||||   :::
Db        895 CGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC---AAT 951

Qy        203 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 222
                  ||||||      |||||||||   :::         :::|||:::||||||
Db        952 GCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGTGGGACA 1011

Qy        223 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
                ||| |||||||||||  ||||||||||||||||||:::           :::
Db       1012 GGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAGAAGGAT 1071

Qy        243 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 262
                |||      :::::::::||||||||||||   |||   |||:::   ||||||
Db       1072 GATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCTTACGGC 1131

Qy        263 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 282
              ||||||::::::||||||   :::|||   |||:::      |||      :::|||
Db       1132 TACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAGGCAGCA 1191

Qy        283 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 302
                |||:::   |||           |||   |||      |||      ::::::
Db       1192 AATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCAGATATT 1248

Qy        303 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 322
              ||||||   :::   |||          ||||||      |||:::|||||   :::|||
Db       1249 TTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCCTTCTCA 1308

Qy        323 PheThr 324
              :::|||
Db       1309 TATACG 1314
```

Search completed: January 2, 2004, 10:05:11
Job time : 845.067 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:47:17 ; Search time 2994.02 Seconds
(without alignments)
2743.777 Million cell updates/sec

Title: US-09-980-881A-3
Perfect score: 1798
Sequence: 1 FQSGQVLAALPRTSRQVQVL.....IKYSFTSNPPVEKLLPLSLK 338

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135916_26873/app_query.fasta_1.1628
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_5167_@runat_31122003_135916_26873 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1494	83.1	1446	11	AK004045	AK004045 Mus muscu
2		1253.5	69.7	971	9	AL536069	AL536069 AL536069
3		988	54.9	853	12	BI144644	BI144644 602909932
4		978	54.4	852	10	BG569281	BG569281 602588573
5		945.5	52.6	1004	14	BY704293	BY704293 BY704293
6		926.5	51.5	738	10	BG616456	BG616456 602614618
7		924	51.4	1022	12	BI332659	BI332659 602984408
8		921	51.2	788	10	BG618629	BG618629 602646064
9		912	50.7	747	10	BG565100	BG565100 602583720
10		903	50.2	702	2	HSM089979	Bx496754 Homo sapi
11		897	49.9	747	10	BG566051	BG566051 602582652
12		894.5	49.7	724	9	AI048167	AI048167 ud71b12.y
13		863	48.0	735	9	AV646979	AV646979 AV646979
14		862	47.9	725	10	BG618239	BG618239 602645967
15		831.5	46.2	688	9	AV651709	AV651709 AV651709
16		828.5	46.1	806	10	BG567264	BG567264 602589745
17		827.5	46.0	577	14	CB154355	CB154355 K-EST0212
18		818	45.5	541	14	CB157619	CB157619 K-EST0216
19		814.5	45.3	750	10	BG618813	BG618813 602646186
20	c	812.5	45.2	973	9	AL536068	AL536068 AL536068
21		806.5	44.9	661	9	AI529872	AI529872 ui83d11.y
22		806	44.8	580	9	AI182113	AI182113 ud73f10.y
23		806	44.8	589	9	AA968307	AA968307 uc71h07.y
24		795	44.2	510	14	CB156851	CB156851 K-EST0215
25		785	43.7	691	10	BG568240	BG568240 602587149
26		776.5	43.2	603	10	AW950700	AW950700 EST362770
27		766	42.6	511	9	AI574117	AI574117 uj68d09.y
28		759	42.2	511	9	AI528146	AI528146 ui94a04.y
29		755	42.0	583	9	AV693037	AV693037 AV693037
30		752	41.8	485	13	BX091627	BX091627 BX091627
31	c	751.5	41.8	843	10	BG216526	BG216526 RST36106
32		744	41.4	602	9	AV692032	AV692032 AV692032
33		741	41.2	601	9	AV658390	AV658390 AV658390
34		735	40.9	623	9	AV694058	AV694058 AV694058
35		732.5	40.7	537	14	N98450	N98450 za28c09.r1
36		732	40.7	434	14	CB164014	CB164014 K-EST0225
37		727.5	40.5	749	10	BG617911	BG617911 602645566
38		725	40.3	621	9	AV698425	AV698425 AV698425
39		715.5	39.8	958	9	AI322676	AI322676 mj24d06.y
40		712	39.6	787	14	CB593837	CB593837 AGENCOURT
41		707	39.3	610	9	AI574064	AI574064 uj67g11.y
42	c	698.5	38.8	723	9	AW413090	AW413090 uq52c03.x
43		688	38.3	487	9	AI574057	AI574057 uj67g04.y
44	c	685.5	38.1	772	9	AI255929	AI255929 ui83d11.x
45		676.5	37.6	692	10	BF384322	BF384322 602046670

ALIGNMENTS

RESULT 1
AK004045
LOCUS AK004045 1446 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1110032P04 product:carboxypeptidase B2 (plasma), full insert sequence.
ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1446)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1446
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1110032P04"
/db_xref="MGI:1900711"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
CDS 42. .1310
/note="unnamed protein product; carboxypeptidase B2 (plasma) (MGD|MGI:1891837, GB|NM_019775, evidence: BLASTN, 99%, match=1419)
putative"
/codon_start=1
/protein_id="BAB23141.1"
/db_xref="GI:12835068"
/db_xref="MGI:1891837"
/translation="MKLHGLGILVAIILYEQHGPAFQSGQVLSALPRTSRQVQLLQNL
TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT 421 a 329 c 311 g 385 t
ORIGIN

Alignment Scores:

Pred. No.:	4.86e-159	Length:	1446
Score:	1494.00	Matches:	292
Percent Similarity:	79.64%	Conservative:	21
Best Local Similarity:	74.30%	Mismatches:	25
Query Match:	83.09%	Indels:	55
DB:	11	Gaps:	2

US-09-980-881A-3 (1-338) x AK004045 (1-1446)

```
Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||:::|||||||||||||||||||||||||||:::|||
Db        105 TTTCAGAGTGGCCAGGTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 164

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||:::|||||||||||||||||||||||:::   |||
Db        165 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 224

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                 |||||||:::|||||||||||||||||||||||||||||:::|||||||||
Db        225 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 284
```

```
Qy          61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
               ||||||||||||||||  ||||||  :::||||||:::   :::||||||||||||||||
Db         285 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 344

Qy          81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
               :::||||||     ||||||||||||||||||||||||||||||||||||||||||||||
Db         345 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 404

Qy         101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
               ||||||||||||||||||||||||||||||||||   |||||||||:::||||||||||
Db         405 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 464

Qy         121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
               ||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||||
Db         465 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 524

Qy         141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
               ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db         525 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 584

Qy         161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
               ||||||||||||||||||||||||||||||||||||||||||:::
Db         585 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 644

Qy         175 ------------------------------------------------------------ 175

Db         645 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 704

Qy         176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                   ||||||||||||||||||||||
Db         705 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAACCGC 764

Qy         184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
               |||   :::   ||||||   |||:::||||||||||||||||||||||||||||||||
Db         765 TCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCCAAA 824

Qy         204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
               ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||||
Db         825 CACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGACTTTAT 884

Qy         224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsnGln 243
               ||||||||||||||||||||||||||||||   ||||||||||||||||:::
Db         885 CCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATCGACCAC 944

Qy         244 IleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProProTyrSerTyr 263
               ||||||||||||||||||||||||||||||||||   |||:::|||||||||||||||||
Db         945 ATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTATTCCTAT 1004

Qy         264 ThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaValArg 283
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1005 AACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCAGTTCGT 1064

Qy         284 AlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThrLeu 303
               |||||||||      :::||||||||||||||||||||||||   |||||||||:::|||
Db        1065 GCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAAAGTTTA 1124

Qy         304 TyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSerPhe 323
               ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        1125 TATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATATTCGTTT 1184

Qy         324 ----------------------------------------------------ThrSerAs 326
                                                                   ||||||||
Db        1185 ACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATACATCAAA 1244

Qy         326 nProProValGluLysLeuLeuProLeuSerLeuLys 338
               ||||   |||:::||||||   |||   |||||||||
Db        1245 CCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 1281
```

RESULT 2
AL536069
LOCUS AL536069 971 bp mRNA linear EST 31-MAY-2003
DEFINITION AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone CS0DF022YE21 5-PRIME, mRNA sequence.

ACCESSION AL536069
VERSION AL536069.2 GI:31260939
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 971)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 13, 2001 this sequence version replaced gi:12799562.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6944.r For more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF022AC11QP1&cluster=6944.r. Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DF022AC11QP1.
FEATURES Location/Qualifiers
source 1. .971
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF022YE21"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."
BASE COUNT 282 a 201 c 211 g 272 t 5 others
ORIGIN

Alignment Scores:

Pred. No.:	6.72e-132	Length:	971
Score:	1253.50	Matches:	240
Percent Similarity:	86.02%	Conservative:	0
Best Local Similarity:	86.02%	Mismatches:	2
Query Match:	69.72%	Indels:	37
DB:	9	Gaps:	1

US-09-980-881A-3 (1-338) x AL536069 (1-971)

```
Qy           1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         131 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 190

Qy          21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         191 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 250

Qy          41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         251 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 310

Qy          61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         311 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 370

Qy          81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         371 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 430

Qy         101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        431 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 490

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        491 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 550

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        551 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 610

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||||
Db        611 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 670

Qy        175 ------------------------------------------------------------ 175

Db        671 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCGGTG 730

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||
Db        731 GTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGAATGTGGAGAAAGAACCGT 790

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        791 TCTTTCTATGCGAACAATCATTGCATTGGAACAGACCTGAATAGGAACTTTGCTTCCAAA 850

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 CACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTACTGTGGACTTTAT 910

Qy        224 ProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 242
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        911 CCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAGAAAWATMAAC 967

RESULT 3
BI144644
LOCUS       BI144644                853 bp    mRNA    linear   EST 05-JUL-2001
DEFINITION  602909932F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5051358 5',
            mRNA sequence.
ACCESSION   BI144644
VERSION     BI144644.1  GI:14604645
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 853)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11139  row: o  column: 07
            High quality sequence stop: 849.
FEATURES             Location/Qualifiers
     source          1. .853
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5051358"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                                Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      243 a    199 c    194 g    217 t
ORIGIN

Alignment Scores:
Pred. No.:              9.96e-102       Length:         853
Score:                  988.00          Matches:        194
Percent Similarity:     78.28%          Conservative:   15
Best Local Similarity:  72.66%          Mismatches:     18
Query Match:            54.95%          Indels:         40
DB:                     12              Gaps:           2

US-09-980-881A-3 (1-338) x BI144644 (1-853)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||:::|||||||||||||||||||||||||||||:::|||
Db         59 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 118

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||:::|||||||||||||||||||||||:::   |||
Db        119 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 178

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                 |||||||:::|||||||||||||||||||||||||||||||:::|||||||||
Db        179 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 238

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||   ||||||   :::||||||:::   :::|||||||||||||||
Db        239 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 298

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::||||||      ||||||||||||||||||||||||||||||||||||||||||||
Db        299 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 358

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||   |||||||||:::||||||||
Db        359 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 418

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        419 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 478

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        479 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 538

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||||:::
Db        539 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 598

Qy        175 ------------------------------------------------------------ 175

Db        599 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCCCGTG 658

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                  |||||||||||||||||||||
Db        659 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAACCGT 718

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
                                  |||:::|||   ||||||||||||||||||||||||
Db        719 ---CTGTCACAAGAACAACCGTGCGTTGGGCCAGACCTGAACAGGAACTTCGCTTCCAAA 775

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGlu-ThrTyrCys-GlyLeuT 223
              ||||||||||||:::|||||||||||||||||||||||||| |||||||| ||||||| 
Db        776 CACTGGTGTGAGAAAGGTGCGTCCAGTTCCTCCTGCTCTGAAAACCTACTGTGGGACTTT 835

Qy        223 yrProGluSerGluPro 228
              |||||||||||||||
Db        836 ATCCTGAGTCTGAGCCG 852

RESULT 4
BG569281
LOCUS       BG569281                852 bp    mRNA    linear   EST 10-APR-2001
```

```
DEFINITION  602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5',
            mRNA sequence.
ACCESSION   BG569281
VERSION     BG569281.1  GI:13576934
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1580  row: b  column: 19
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4722354"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      251 a    177 c    187 g    237 t
ORIGIN

Alignment Scores:
Pred. No.:              1.37e-100       Length:         852
Score:                  978.00          Matches:        197
Percent Similarity:     76.15%          Conservative:   1
Best Local Similarity:  75.77%          Mismatches:     5
Query Match:            54.39%          Indels:         58
DB:                     10              Gaps:           2

US-09-980-881A-3 (1-338) x BG569281 (1-852)

Qy        135 ProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIle 154
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db          2 CCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATT 61

Qy        155 AspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGly 174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGC 121

Qy        175 His------------------------------------------------------- 175
              |||
Db        122 CATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGAT 181

Qy        176 ----------------------------------------------------AsnArg 177
                                                                   ||||||
Db        182 TTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGA 241

Qy        178 MetTrp-ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAs 197
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||

Db        242 ATGTGGACGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAA 301

Qy        197 nArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGl 217
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGA 361

Qy        217 uThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLe 237
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTT 421

Qy        237 uArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIl 257
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATAT 481

Qy        257 eValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVa 277
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGT 541

Qy        277 lAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGl 297
              |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db        542 AGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGC 601

Qy        297 yHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr-AspL 317
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||     |
Db        602 GCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGCATT 661

Qy        317 euGlyIleLysTyrSerPhe---------------------------------------- 323
              ||||||||:::|||||||||
Db        662 TGGGCATCAGATATTCGTTTACAATTGAACCTTCGAGATACGGGCACATAACGGATTCTT 721

Qy        324 --------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 337
                           ||||||   ||||||||||||||||||||||||||||||||
Db        722 GCTGCCGGAGCGTTACATCAAA-CCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTA 776

RESULT 5
BY704293
LOCUS       BY704293                1004 bp    mRNA    linear   EST 16-DEC-2002
DEFINITION  BY704293 RIKEN full-length enriched, 18-day embryo whole body Mus
            musculus cDNA clone 1110032P04 5', mRNA sequence.
ACCESSION   BY704293
VERSION     BY704293.1  GI:27115417
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1004)
  AUTHORS   Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
```

```
  TITLE     Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
  JOURNAL   Nature 420, 563-573 (2002)
  MEDLINE   22354683
   PUBMED   12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda
            ,S., Hashizume,W., Hayashida,K., Hirozane,T., Hori,F., Imotani,K.,
            Ishii,Y., Itoh,M., Kagawa,I., Kawai,J., Kojima,Y., Kondo,S., Konno
            ,H., Koya,S., Miyazaki,A., Murata,M., Nakamura,M., Nomura,K.,
            Numazaki,R., Ohno,M., Ohsato,N., Saito,R., Sakazume,N., Sano,H.,
            Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami,M., Takeda,Y.,
            Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
             Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1. .1004
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="1110032P04"
                     /tissue_type="whole body"
                     /dev_stage="18-day embryo"
                     /clone_lib="RIKEN full-length enriched, 18-day embryo
                     whole body"
BASE COUNT      286 a    228 c    228 g    262 t
ORIGIN

Alignment Scores:
Pred. No.:              8.95e-97       Length:        1004
Score:                  945.50         Matches:       188
Percent Similarity:     76.69%         Conservative:  16
Best Local Similarity:  70.68%         Mismatches:    24
Query Match:            52.59%         Indels:        41
DB:                     14             Gaps:          2

US-09-980-881A-3 (1-338) x BY704293 (1-1004)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||:::||||||||||||||||||||||||||||:::|||
Db        105 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 164

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||:::|||||||||||||||||||||||::   |||
Db        165 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 224

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                 |||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db        225 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 284

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||   ||||||   :::||||||:::   :::|||||||||||||||
Db        285 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 344

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::||||||      ||||||||||||||||||||||||||||||||||||||||||||
Db        345 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 404

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||   |||||||||:::|||||||||
Db        405 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 464

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||||
Db        465 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 524

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||      ||||||||||||||||||||||||||||||||||||
Db        525 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 584

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175
              ||||||||||||||||||||||||||||||||||||||||:::
Db        585 AGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 644

Qy        175 ------------------------------------------------------------ 175

Db        645 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGATTTCTACATCATGCGCGTG 704

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||   |||
Db        705 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGTATGTGGAGGAAGAAGCGC 764

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::   ||||||   |||:::||||||||||||||||||||||||||||||||
Db        765 TCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCCAAA 824

Qy        204 HisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeuTyr 223
              ||||||||||||:::                ||||||   ||||||||||   |||
Db        825 CACTGGTGTGAGAAG---GTGCGTAAGTTCCTCTGCTCTGA-ACCTACTGTGGATT-TAT 879

Qy        224 ProGluSerGluProGlu 229
                  ||||||||||||
Db        880 TCTAG-TCTGAGCCAGAG 896

RESULT 6
BG616456
LOCUS       BG616456                738 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602614618F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4733500 5',
            mRNA sequence.
ACCESSION   BG616456
VERSION     BG616456.1  GI:13667827
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 738)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1595  row: c  column: 05
```

```
          High quality sequence stop: 666.
FEATURES             Location/Qualifiers
     source          1. .738
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4733500"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      209 a    161 c    153 g    215 t
ORIGIN

Alignment Scores:
Pred. No.:              8.12e-95       Length:         738
Score:                  926.50         Matches:        192
Percent Similarity:     71.13%         Conservative:   10
Best Local Similarity:  67.61%         Mismatches:     27
Query Match:            51.53%         Indels:         55
DB:                     10             Gaps:           3

US-09-980-881A-3 (1-338) x BG616456 (1-738)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         40 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 99

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        100 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 159

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        160 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 219

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        220 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 279

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        280 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 339

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        340 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 399

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        400 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 459

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        460 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 519

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--AsnArgMetTrpA 180
              ||||||||||||||||||||||||||||||||||||||||||||  |||   :::
Db        520 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 579

Qy        180 rgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnP 200
                   ||||||:::   |||   :::         |||
Db        580 GGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGA------------------- 620

Qy        200 heAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrC 220
                                 |||      |||   |||                      |
Db        621 ----------TTTCTATGTTATGCCGGTGGTTCA----------------------T 645

Qy        220 ysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgA 240
              ||||||||
Db        646 GTGGACTG---------------------------------------------------- 653

Qy        240 snIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheP 260
                                                                        |||
Db        654 ----------------------------------------------------------GTTATGG 660

Qy        260 roTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerG 280
              ||||||:::|||:::::::::::::         :::|||      |||          |
Db        661 ACTACTCATGGACCAAGAATCGCACTGTTGGACAAAAGGAACCCGGTCTTTACTAATGCG 720

Qy        280 luAlaVal 282
              ||   :::
Db        721 AAACAATT 728

RESULT 7
BI332659
LOCUS       BI332659                1022 bp    mRNA    linear   EST 30-JUL-2001
DEFINITION  602984408F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5137369 5',
            mRNA sequence.
ACCESSION   BI332659
VERSION     BI332659.1  GI:15017316
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1022)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11337  row: o  column: 02
            High quality sequence stop: 841.
FEATURES             Location/Qualifiers
     source          1. .1022
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5137369"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      303 a    233 c    225 g    261 t
ORIGIN

Alignment Scores:
Pred. No.:              2.59e-94       Length:         1022
Score:                  924.00         Matches:        200
Percent Similarity:     73.67%         Conservative:   21
Best Local Similarity:  66.67%         Mismatches:     37
Query Match:            51.39%         Indels:         45
DB:                     12             Gaps:           2

US-09-980-881A-3 (1-338) x BI332659 (1-1022)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
```

```
Db        60 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 119

Qy        21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Db       120 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 179

Qy        41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Db       180 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 239

Qy        61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Db       240 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 299

Qy        81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Db       300 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 359

Qy       101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Db       360 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 419

Qy       121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Db       420 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 479

Qy       141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160

Db       480 GTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 539

Qy       161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--------------- 175

Db       540 AGAGAATGGATTTCACCTGCTTTCTGTTG-TGGTTCATAGGCTACGTGACACAATTCCAT 598

Qy       175 ------------------------------------------------------------ 175

Db       599 GGGAAAGAAAATCTGTATACCAGACTTCTGAGGCACGTGGAATTTCTACATCATGCCCGT 658

Qy       176 ----------------------------------------AsnArgMetTrpArgLysAsn 182

Db       659 GATGGAACGTGGATGGCTATGACTACACGTGGACAAAAGAATCGAATGTTGCAGGAAGAA 718

Qy       183 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThr---AspLeuAsnArgAsnPheAla 201

Db       719 CCGTCTGCTCACCAGAACAACCGCTGCTGTGGGCAACATGACCTGAACAGGAACTTCGCT 778

Qy       202 SerLysHisTrpCysGluGluGly-AlaSerSerSerSerCysSerGluThrTyrCysGl 221

Db       779 TCCAAACACTGGTGTGAGAAAGGTTGCGTCAAGTTTCCTCCTGTCTGAAAC-TACTGTGG 837

Qy       221 yLeuTyr-ProGluSer-GluProGluValLysAlaValAlaSerPheLeuArgArgAsn 240

Db       838 ACTTTATTCCCGAGTCTAGAGCCAGAAGTGAAGGCCAGGGGTGACTTTTTGAGAAGAATA 897

Qy       241 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleVal 258

Db       898 TCTGACCCCCTTAAAGT-TACTTCAAGATGGCTCATAATTCCACAAATATGGTT 950
```

RESULT 8
BG618629

```
LOCUS       BG618629                788 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602646064F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767974 5',
            mRNA sequence.
ACCESSION   BG618629
VERSION     BG618629.1  GI:13670000
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 788)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1630  row: o  column: 15
            High quality sequence stop: 760.
FEATURES             Location/Qualifiers
     source          1. .788
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767974"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      222 a    157 c    169 g    240 t
ORIGIN
```

Alignment Scores:

Pred. No.:	3.81e-94	Length:	788
Score:	921.00	Matches:	179
Percent Similarity:	82.19%	Conservative:	1
Best Local Similarity:	81.74%	Mismatches:	1
Query Match:	51.22%	Indels:	38
DB:	10	Gaps:	1

US-09-980-881A-3 (1-338) x BG618629 (1-788)

```
Qy         1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20

Db       132 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 191

Qy        21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40

Db       192 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 251

Qy        41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60

Db       252 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 311

Qy        61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80

Db       312 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 371

Qy        81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100

Db       372 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 431

Qy       101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120

Db       432 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 491

Qy       121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140

Db       492 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 551

Qy       141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
```

```
                ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||
Db        552 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 611

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
                |||||||||||||||||||||||||||||||||||||||||||
Db        612 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTAT 671

Qy        175 ------------------------------------------------------------ 175

Db        672 GGGATAATAGGGCAATATTACCAATCTCCTGAGGCTTGTGGATTTCTATGTTATGCCAGT 731

Qy        176 ------------------------------------AsnArgMetTrpArgLys 181
                                                   ||||||:::|||||||||
Db        732 GGTTAATGTTGGATGGGTTATGACTACTCATGGAAAAGGAATCGATTGTGGAGAAAG 788

RESULT 9
BG565100
LOCUS       BG565100                747 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602583720F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4711348 5',
            mRNA sequence.
ACCESSION   BG565100
VERSION     BG565100.1  GI:13572753
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1551  row: h  column: 05
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4711348"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      220 a    154 c    166 g    207 t
ORIGIN

Alignment Scores:
Pred. No.:              3.72e-93        Length:         747
Score:                  912.00          Matches:        196
Percent Similarity:     80.48%          Conservative:   6
Best Local Similarity:  78.09%          Mismatches:     7
Query Match:            50.72%          Indels:         45
DB:                     10              Gaps:           2

US-09-980-881A-3 (1-338) x BG565100 (1-747)
```

```
Qy         30 ValLeuTrpGlnProValThrAlaAspLeuIleValLysLysLysGlnValHisPhePhe 49
              ||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db          1 GTTCTCTGGCAGCCGGTAACAGCTGACCTTAT-GTGAAGAAAAAACAAGTCCATTTTTTT 59

Qy         50 ValAsnAlaSerAspValAspAsnValLysAlaHisLeuAsnValSerGlyIleProCys 69
              |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 GTAGATGCATCTGATGTCGACAATGTGAAAGCCCATTTAAATGTGAGCGGAATTCCATGC 119

Qy         70 SerValLeuLeuAlaAspValGluAspLeuIleGlnGlnGlnIleSerAsnAspThrVal 89
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        120 AGTGTCTTGCTGGCAGATGTGGAAGATCTTATTCAACAGCAGATTTCCAACGACACAGTC 179

Qy         90 SerProArgAlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSer 109
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        180 AGCCCCCGAGCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCT 239

Qy        110 TrpIleGluPheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySer 129
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        240 TGGATAGAATTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCC 299

Qy        130 SerPheGluLysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLys 149
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        300 TCATTTGAGAAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAA 359

Qy        150 AsnAlaIleTrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCys 169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 AATGCCATATGGATTGACTGTGGAATCCATGCCAGAGAATGGACCTCTCCTGCTTTCTGC 419

Qy        170 LeuTrpPheIleGlyHis---------------------------------------------- 175
              ||||||||||||||||
Db        420 TTGTGGTTCATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTC 479

Qy        176 ----------------------------------------------------AsnArgMet 178
                                                                   :::
Db        480 CTGAGGCTTGTGGATTTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCA 539

Qy        179 TrpArg--------------------LysAsnArgSerPheTyrAlaAsnAsnHisCys 191
              |||:::                    :::||||||||||||||||||||||||||||||
Db        540 TGGAAACAAGAATCGGAATGTGGAGAACAGAACCGTTCTTTCTATGCGAACAATCATTGC 599

Qy        192 IleGlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGlu-GlyAlaSe 211
              ||||||||||||||||||||||  ||||||||||||||||||||||||| ||||||||
Db        600 ATCGGAACAGACCTGAATAGGAACTT-GCTTCCAAACACTGGTGTGAGGAAGGGTGCATC 658

Qy        211 rSerSerSer-CysSerGlu-ThrTyr-CysGlyLeuTyrProGluSerGluProGluVa 230
              |||||||||| ||||||||| |||||| ||||||||||||||||||||||||||||||
Db        659 CAGTTCCTCATTGCTCGGAAAACCTACTTGTGGACTTTATCCTGAGTCAGAACCAGAGTG 718

Qy        230 lLysAlaValAlaSerPheLeuArg 238
               :::||||||||||||||||:::
Db        719 -GAAGCAGTGGCTAGTTTCTTGAAG 742

RESULT 10
HSM089979
ID   HSM089979  standard; RNA; EST; 702 BP.
XX
AC   BX496754;
XX
SV   BX496754.1
XX
DT   09-MAY-2003 (Rel. 75, Created)
DT   09-MAY-2003 (Rel. 75, Last updated, Version 1)
XX
DE   Homo sapiens mRNA; EST DKFZp779I1529_r1  (from clone DKFZp779I1529)
XX
KW   EST; expressed sequence tag.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-702
```

```
RA   Ansorge W., Krieger S., Regiert T., Rittmueller C., Schwager B.,
RA   Mewes H.W., Weil B., Amid C., Osanger A., Fobo G., Han M., Wiemann S.;
RT   ;
RL   Submitted (07-MAY-2003) to the EMBL/GenBank/DDBJ databases.
RL   MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
XX
CC   This is the 5' sequence of the clone insert
CC   Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC   Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
CC   sequenced by EMBL (European Molecular Biology Laboratories,
CC   Heidelberg/Germany) within the cDNA sequencing consortium of
CC   the German Genome Project.
CC   No s1 sequence available.
CC   This clone (DKFZp779I1529) is available at the RZPD in Berlin.
CC   Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC   14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .702
FT                   /db_xref="taxon:9606"
FT                   /mol_type="mRNA"
FT                   /organism="Homo sapiens"
FT                   /clone="DKFZp779I1529"
FT                   /clone_lib="779 (synonym: hncc1). Vector pSport1_Sfi; host
FT                   DH10B; sites SfiIA + SfiIB"
FT                   /dev_stage="fetal"
FT                   /tissue_type="liver"
XX
SQ   Sequence 702 BP; 197 A; 149 C; 146 G; 210 T; 0 other;

Alignment Scores:
Pred. No.:              3.58e-92        Length:         702
Score:                  903.00          Matches:        174
Percent Similarity:     98.86%          Conservative:   0
Best Local Similarity:  98.86%          Mismatches:     1
Query Match:            50.22%          Indels:         1
DB:                     2               Gaps:           0

US-09-980-881A-3 (1-338) x HSM089979 (1-702)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 184

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 244

Qy         41 ValLys-LysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAl 60
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        245 GTGAAGAAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGC 304

Qy         60 aHisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIl 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        305 CCATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTAT 364

Qy         80 eGlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGl 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        365 TCAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACA 424

Qy        100 nTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAs 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        425 GTATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGA 484

Qy        120 pMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLy 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        485 TATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAA 544

Qy        140 sValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAl 160
              |||||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db        545 GGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGC 604

Qy        160 aArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        605 CAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 650

RESULT 11
BG566051
LOCUS       BG566051                747 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602582652F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4710564 5',
            mRNA sequence.
ACCESSION   BG566051
VERSION     BG566051.1  GI:13573704
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1549  row: g  column: 13
            High quality sequence stop: 731.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4710564"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      212 a    157 c    159 g    219 t
ORIGIN

Alignment Scores:
Pred. No.:              1.91e-91        Length:         747
Score:                  897.00          Matches:        178
Percent Similarity:     92.35%          Conservative:   3
Best Local Similarity:  90.82%          Mismatches:     12
Query Match:            49.89%          Indels:         3
DB:                     10              Gaps:           0

US-09-980-881A-3 (1-338) x BG566051 (1-747)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||
Db         86 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTATAACCTCTAGGCAAGTTCAAGTTCTA 145

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        146 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGCTGACCTTATT 205

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        206 GTGAAGAAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 265
```

```
Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        266 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 325

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        326 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 385

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        386 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 445

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrVal-LeuLy 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        446 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTCTAAA 505

Qy        140 sValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAl 160
              |||||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db        506 GGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGC 565

Qy        160 aArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--AsnArgMetTrp 179
              |||||||||||||||||||||||||||||||||||||||||||||  |||  :::
Db        566 CAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATAACTCAATTCTA 625

Qy        180 ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGly 193
                 ||||||:::   |||   :::          |||
Db        626 TGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGA 667

RESULT 12
AI048167
LOCUS       AI048167                724 bp    mRNA    linear   EST 08-JUL-1998
DEFINITION  ud71b12.y1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1451327 5' similar to TR:Q15114 Q15114 PREPRO-PLASMA
            CARBOXYPEPTIDASE B. ;, mRNA sequence.
ACCESSION   AI048167
VERSION     AI048167.1  GI:3296454
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 724)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:924643
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .724
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1451327"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse liver mlia"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      215 a    169 c    160 g    178 t      2 others
ORIGIN

Alignment Scores:
Pred. No.:              3.5e-91         Length:         724
Score:                  894.50          Matches:        171
Percent Similarity:     78.42%          Conservative:   18
Best Local Similarity:  70.95%          Mismatches:     15
Query Match:            49.75%          Indels:         37
DB:                     9               Gaps:           1

US-09-980-881A-3 (1-338) x AI048167 (1-724)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||:::|||||||||||||||||||||||||||:::|||
Db          2 TTTCAGAGTGGCCAGGTTTTATCTGCTCTTCCAAGAACCTCCAGGCAAGTTCAACTACTT 61

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||:::|||||||||||||||||||||||:::   |||
Db         62 CAGAATCTTACTACAACGTATGAGGTCGTTCTCTGGCAGCCAGTGACAGCTGAATTCATC 121

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
                 |||||||:::|||||||||||||||||||||||||||||||:::||||||||||
Db        122 GAGAAGAAAAAGGAAGTCCACTTTTTTGTGAATGCGTCTGATGTCGACAGTGTCAAAGCG 181

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              |||||||||||||||   ||||||   :::||||||:::   :::|||||||||||||||
Db        182 CATTTAAATGTGAGCAGAATTCCATTTAACGTTCTGATGAACAACGTGGAGGACCTAATT 241

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              :::||||||      ||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GAACAGCAGACTTTCAATGACACGGTCAGCCCCCGCGCCTCCGCTTCATACTATGAGCAG 301

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TATCACTCGCTAAATGAAATCTATTCCTGGATAGAAGTCATAACTGAACAGCATCCTGAC 361

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db        362 ATGCTCCAGAAAATCTACATCGGATCATCATTCGAGAAGTACCCACTTTATGTTTTAAAG 421

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              |||||||||:::||||||      ||||||||||||||||||||||||||||||||||||||
Db        422 GTCTCAGGACAGGAACAAAGAATCAAAAATGCCATCTGGATCGACTGTGGAATCCATGCC 481

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis---------------- 175
              ||||||||||||||||||   ||||||||||||||||:::
Db        482 AGAGAATGGATTTCACCTGCTCTCTGTTTGTGGTTCATAGGCTACGTGACACAATTCCAT 541

Qy        175 ------------------------------------------------------------ 175

Db        542 GGGAGAGAANATCTGTATACCAGACTTCTGAGGCACGTGGATCTCTACATCATGCACGTG 601

Qy        176 ----------------------------------AsnArgMetTrpArgLysAsnArg 183
                                                 ||||||||||||||||:::|||
Db        602 ATGAACGTGGATGGCTATGACTACACGTGGAAAAAGAATCGAATGTGGAGGAAGAGCCGN 661

Qy        184 SerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSerLys 203
              |||   :::   |||:::   |||:::||||||||||||||||||||||||||||||||
```

```
Db        662 TCTGCTCACAAGAACCACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCCAAA 721

Qy        204 His 204
                |||
Db        722 CAC 724

RESULT 13
AV646979
LOCUS       AV646979                735 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV646979 GLC Homo sapiens cDNA clone GLCATD06 3', mRNA sequence.
ACCESSION   AV646979
VERSION     AV646979.1  GI:9867993
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCATD06"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      200 a    158 c    172 g    205 t
ORIGIN

Alignment Scores:
Pred. No.:              1.39e-87       Length:          735
Score:                  863.00         Matches:         170
Percent Similarity:     96.59%         Conservative:    0
Best Local Similarity:  96.59%         Mismatches:      5
Query Match:            48.00%         Indels:          1
DB:                     9              Gaps:            0

US-09-980-881A-3 (1-338) x AV646979 (1-735)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         58 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 117

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        118 CAGAATCTTACTACAACATATGAGATTGTTCTCTGGCAGCCGGTAACAGGTGACCTTATT 177

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        178 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 237

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        238 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGACGTGGAAGATCTTATT 297

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db        298 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGGACTATGAACAG 357

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        358 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 417

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        418 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 477

Qy        141 -ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAl 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        478 GGTGTCTGGAAAAGAACAAACAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGC 537

Qy        160 aArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
              ||||||||||||||||||||||||    |||       |||||||||
Db        538 CAGAGAATGGATCTCTCCTGCTTTCTGGTTGGGGGTCATAGGCCAT 583

RESULT 14
BG618239
LOCUS       BG618239                725 bp    mRNA    linear   EST 18-APR-2001
DEFINITION  602645967F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4767376 5',
            mRNA sequence.
ACCESSION   BG618239
VERSION     BG618239.1  GI:13669610
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1629  row: f  column: 17
            High quality sequence stop: 723.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4767376"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      203 a    153 c    149 g    220 t
ORIGIN

Alignment Scores:
```

```
Pred. No.:              1.77e-87        Length:         725
Score:                  862.00          Matches:        172
Percent Similarity:     98.29%          Conservative:   0
Best Local Similarity:  98.29%          Mismatches:     3
Query Match:            47.94%          Indels:         2
DB:                     10              Gaps:           0

US-09-980-881A-3 (1-338) x BG618239 (1-725)

Qy          1 PheGlnSerGlyGlnValLeuAlaAlaLeuProArgThrSerArgGlnValGlnValLeu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 TTTCAGAGTGGCCAAGTTCTAGCTGCTCTTCCTAGAACCTCTAGGCAAGTTCAAGTTCTA 190

Qy         21 GlnAsnLeuThrThrThrTyrGluIleValLeuTrpGlnProValThrAlaAspLeuIle 40
              |||||||||||||||||||||||   |||||||||||||||||||||||||||||||
Db        191 CAGAATCTTACTACAACATATGAGAT-GTTCTCTGGCAGCCGGTAACAGCTGACCTTAT- 248

Qy         41 ValLysLysLysGlnValHisPhePheValAsnAlaSerAspValAspAsnValLysAla 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        249 GTGAAGAAAAAACAAGTCCATTTTTTTGTAAATGCATCTGATGTCGACAATGTGAAAGCC 308

Qy         61 HisLeuAsnValSerGlyIleProCysSerValLeuLeuAlaAspValGluAspLeuIle 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        309 CATTTAAATGTGAGCGGAATTCCATGCAGTGTCTTGCTGGCAGATGTGGAAGATCTTATT 368

Qy         81 GlnGlnGlnIleSerAsnAspThrValSerProArgAlaSerAlaSerTyrTyrGluGln 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        369 CAACAGCAGATTTCCAACGACACAGTCAGCCCCCGAGCCTCCGCATCGTACTATGAACAG 428

Qy        101 TyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArgHisProAsp 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        429 TATCACTCACTAAATGAAATCTATTCTTGGATAGAATTTATAACTGAGAGGCATCCTGAT 488

Qy        121 MetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyrValLeuLys 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        489 ATGCTTACAAAAATCCACATTGGATCCTCATTTGAGAAGTACCCACTCTATGTTTTAAAG 548

Qy        141 ValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGlyIleHisAla 160
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        549 GTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATTGACTGTGGAATCCATGCC 608

Qy        161 ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 175
              |||||||||||||||||||||||||||||||||||||||||||||
Db        609 AGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCAT 653

RESULT 15
AV651709
LOCUS       AV651709                688 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV651709 GLC Homo sapiens cDNA clone GLCCSF10 3', mRNA sequence.
ACCESSION   AV651709
VERSION     AV651709.1  GI:9872723
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCCSF10"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      208 a    145 c    151 g    184 t
ORIGIN

Alignment Scores:
Pred. No.:              4.89e-84        Length:         688
Score:                  831.50          Matches:        162
Percent Similarity:     89.50%          Conservative:   0
Best Local Similarity:  89.50%          Mismatches:     1
Query Match:            46.25%          Indels:         18
DB:                     9               Gaps:           1

US-09-980-881A-3 (1-338) x AV651709 (1-688)

Qy        176 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 105

Qy        196 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 215
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 165

Qy        216 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 235
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 225

Qy        236 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 285

Qy        256 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 275
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 345

Qy        276 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 295
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 405

Qy        296 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 465

Qy        316 AspLeuGlyIleLysTyrSerPhe---------------------------------- 323
              |||||||||||||||||||||||
Db        466 GATTTGGGCATCAAATATTCGTTTACAATTGAACCTCGAGATACCGGCACATACCGATTC 525

Qy        324 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 338
                            ||||||||||||||||||||||||||||||   ||||||||||
Db        526 TTTGTGGCGGAGCGTTACATCAAACCCACCTGTTGAGAAGCTTTTGGCGCTGTCTCTTAA 585

Qy        338 s 338
              |
Db        586 A 586


Search completed: January  2, 2004, 02:37:30
Job time : 3005.02 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:33:22 ; Search time 268.597 Seconds
(without alignments)
2472.334 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135914_26819/app_query.fasta_1.1628
-DB=N_Geneseq_19Jun03 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_577_@runat_31122003_135914_26819 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1325	99.0	1573	21	AAC81962	Human brain carbox
2	1289.5	96.4	1625	25	ABX13670	Human protease cDN
3	1272	95.1	1272	20	AAV74302	Human plasma carbo
4	1272	95.1	1728	24	ABN95899	Gene #2397 used to
5	1272	95.1	1749	17	AAT11671	Human plasma carbo
6	1272	95.1	1749	18	AAT62846	Human plasma carbo
7	1267	94.7	1749	14	AAQ41001	Human plasma carbo
8	902	67.4	1400	21	AAF18005	Lung cancer associ
9	610	45.6	927	17	AAT35760	Rat mature carboxy
10	582	43.5	921	16	AAQ90601	Porcine carboxypep
11	582	43.5	1215	16	AAQ90600	Porcine Tyr-His-Me
12	567	42.4	999	17	AAT42500	Mature HCPB coding
13	567	42.4	1053	17	AAT42497	mature HCPB-(His)6
14	567	42.4	1263	17	AAT42494	Human pancreatic c
15	567	42.4	1263	19	AAV41795	Human pancreatic c
16	567	42.4	1284	17	AAT42506	ProHCPB gene with
17	560	41.9	1053	20	AAX24804	Human carboxypepta
18	560	41.9	1059	17	AAT42511	Modified HCPB (D25
19	560	41.9	1059	18	AAT62787	Carboxypeptidase B
20	559	41.8	1059	17	AAT42512	Modified HCPB (D25
21	559	41.8	1059	18	AAT62788	Carboxypeptidase B
22	556.5	41.6	1302	25	ABV77254	Nucleotide sequenc
23	556.5	41.6	1311	22	AAD03837	Human carboxypepti
24	556.5	41.6	1993	24	AAD30579	Human protease, PR
25	556	41.6	1059	18	AAT62806	Carboxypeptidase B
26	554	41.4	1059	18	AAT62790	Carboxypeptidase B
27	554	41.4	1059	18	AAT62801	Carboxypeptidase B
28	553	41.3	1059	18	AAT62789	Carboxypeptidase B
29	552	41.3	1059	18	AAT62791	Carboxypeptidase B
30	552	41.3	2023	25	ABZ75116	Anti-human seminal
31	551	41.2	1059	18	AAT62805	Carboxypeptidase B
32	551	41.2	1059	18	AAT62803	Carboxypeptidase B
33	551	41.2	1059	18	AAT62804	Carboxypeptidase B
34	550.5	41.1	1125	22	AAH76477	Human secreted met
35	550.5	41.1	1125	22	AAH76479	cDNA encoding huma
36	550.5	41.1	1125	24	ABK31745	DNA encoding novel
37	550.5	41.1	1222	25	ABT33353	NOVX DNA sequence
38	550.5	41.1	1342	22	AAH76476	Human secreted met
39	550.5	41.1	1342	22	AAH76478	Human secreted met
40	550.5	41.1	1344	25	ABQ77393	Human zinc-binding
41	550.5	41.1	1378	25	ABT33355	NOVX DNA sequence
42	550.5	41.1	1603	24	AAD39053	Human 23566 (carbo
43	550.5	41.1	1826	22	AAF81728	Human protease and
44	549	41.0	1059	18	AAT62808	Carboxypeptidase B
45	549	41.0	1059	18	AAT62800	Carboxypeptidase B

ALIGNMENTS

RESULT 1
AAC81962
ID AAC81962 standard; cDNA; 1573 BP.
XX
AC AAC81962;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B cDNA.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma; ss.
XX
OS Homo sapiens.

```
XX
FH   Key             Location/Qualifiers
FT   CDS             18..1100
FT                   /*tag= a
FT                   /product= "carboxypeptidase B"
XX
PN   WO200066717-A1.
XX
PD   09-NOV-2000.
XX
PF   01-MAY-2000; 2000WO-JP02878.
XX
PR   30-APR-1999;   99JP-0125169.
XX
PA   (MATS/) MATSUMOTO A.
XX
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
DR   P-PSDB; AAB11457.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Claim 2b; Page 64-68; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence 1573 BP; 482 A; 328 C; 294 G; 469 T; 0 other;

Alignment Scores:
Pred. No.:              4.85e-142      Length:        1573
Score:                  1325.00        Matches:       243
Percent Similarity:     99.19%         Conservative:  1
Best Local Similarity:  98.78%         Mismatches:    2
Query Match:            99.03%         Indels:        0
DB:                     21             Gaps:          0

US-09-980-881A-4 (1-246) x AAC81962 (1-1573)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        360 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 419

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        420 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 479

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        480 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 539

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        540 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 599

Qy         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 659

Qy        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
              ||||||||||||||   ||||||  ||||||||||||||||||||||||||||||||||||
Db        660 GGAACAGACCTGAATAGCAACTTTGTCTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 719

Qy        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 779

Qy        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 839

Qy        161 TyrSerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 899

Qy        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        900 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGACAAAACTAGTAAAAATACC 959

Qy        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1019

Qy        221 TrpIleTyrAspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTT 1079

Qy        241 LeuProLeuSerLeuLys 246
              ||||||||||||||||||
Db       1080 TTGCCGCTGTCTCTAAAA 1097

RESULT 2
ABX13670
ID   ABX13670 standard; cDNA; 1625 BP.
XX
AC   ABX13670;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease cDNA.
XX
KW   Human; gene; ss; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..16
FT                   /*tag= a
FT   CDS             17..1177
FT                   /*tag= b
FT                   /product= "Protease"
FT   variation       replace(521,A)
FT                   /*tag= c
FT                   /standard_name= "Single nucleotide polymorphism"
FT   3'UTR           1178..1625
FT                   /*tag= d
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
```

```
DR   WPI; 2003-102387/09.
DR   P-PSDB; ABG72093, ABG72095.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 4; Fig 1; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   cDNA, the gene for which is located on chromosome 13.
XX
SQ   Sequence 1625 BP; 488 A; 334 C; 315 G; 488 T; 0 other;

Alignment Scores:
Pred. No.:              6.15e-138     Length:          1625
Score:                  1289.50       Matches:         244
Percent Similarity:     92.42%        Conservative:    0
Best Local Similarity:  92.42%        Mismatches:      2
Query Match:            96.38%        Indels:          18
DB:                     25            Gaps:            1

US-09-980-881A-4 (1-246) x ABX13670 (1-1625)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 418

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 478

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||
Db        479 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 538

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 598

Qy         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 658

Qy        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GGAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 718

Qy        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 778

Qy        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 838

Qy        161 TyrSerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 898

Qy        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACC 958

Qy        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1018

Qy        221 TrpIleTyrAspLeuGlyIleLysTyrSerPhe-------------------------- 231
              ||||||||||||||||||||||||||||||||
Db       1019 TGGATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACA 1078

Qy        232 ------------------------ThrSerAsnProProValGluLysLeuLeuProLe 243
                                       ||||||||||||||||||||||||||||||||||
Db       1079 TACGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 1138

Qy        243 uSerLeuLys 246
              ||||||||||
Db       1139 GTCTCTAAAA 1148

RESULT 3
AAV74302
ID   AAV74302 standard; cDNA; 1272 BP.
XX
AC   AAV74302;
XX
DT   28-APR-1999  (first entry)
XX
DE   Human plasma carboxypeptidase B (PCPB) thr147 coding sequence.
XX
KW   Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW   polymorphism detection; thrombotic disease; ds.
XX
OS   Homo sapiens.
XX
PN   WO9855645-A1.
XX
PD   10-DEC-1998.
XX
PF   02-JUN-1998;   98WO-EP03244.
XX
PR   03-JUN-1997;   97US-0869057.
XX
PA   (SCHD ) SCHERING AG.
XX
PI   Morser MJ,  Nagashima M;
XX
DR   WPI; 1999-045800/04.
DR   P-PSDB; AAW92270.
XX
PT   Detecting new polymorphism of human plasma carboxypeptidase B -
PT   comprises Alanine or Threonine at position 147 of protein by DNA or
PT   protein analysis, useful to detect risk of thrombotic disease in
PT   humans
XX
PS   Example 1; Page 24; 35pp; English.
XX
CC   This sequence encodes the human plasma carboxypeptidase B (PCPB) mutant
CC   hPCPBthr147. The invention relates to a method for determining the
CC   presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC   comprises obtaining a prepared tissue or blood sample and determining the
CC   presence of DNA coding for naturally occurring polymorphs of the protein
CC   containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
```

```
CC   respectively). Determination of the relative distribution of the PCPB
CC   polymorphs in a patient's blood by genetic or protein analysis by the
CC   methods is useful to determine the risk of thrombotic disease in humans.
CC   Such assessments may be made by accumulating information concerning the
CC   relative distribution of the different polymorphs within the general
CC   population compared with populations known to be at risk and establishing
CC   a PCPB polymorph profile for at-risk patients.
XX
SQ   Sequence 1272 BP; 375 A; 269 C; 271 G; 357 T; 0 other;

Alignment Scores:
Pred. No.:              4.41e-136      Length:        1272
Score:                  1272.00        Matches:       246
Percent Similarity:     81.73%         Conservative:  0
Best Local Similarity:  81.73%         Mismatches:    0
Query Match:            95.07%         Indels:        55
DB:                     20             Gaps:          2

US-09-980-881A-4 (1-246) x AAV74302 (1-1272)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 642

Qy         83 ---------------------------------------------------------- 83

Db        643 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 882

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1062

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
```

```
              ||||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1182

Qy        232 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                              ||||||||||||||||||||||||||||||||||||||||||
Db       1183 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
               |
Db       1243 A 1243

RESULT 4
ABN95899
ID   ABN95899 standard; DNA; 1728 BP.
XX
AC   ABN95899;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #2397 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 2397; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1728 BP; 518 A; 354 C; 338 G; 518 T; 0 other;

Alignment Scores:
Pred. No.:              6.93e-136      Length:        1728
Score:                  1272.00        Matches:       246
Percent Similarity:     81.73%         Conservative:  0
Best Local Similarity:  81.73%         Mismatches:    0
Query Match:            95.07%         Indels:        55
DB:                     24             Gaps:          2
```

```
US-09-980-881A-4 (1-246) x ABN95899 (1-1728)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 421

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 481

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 541

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 601

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        602 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 661

Qy         83 ------------------------------------------------------------ 83

Db        662 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 721

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 781

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 841

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 901

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 961

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1021

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1141

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1142 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1201

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||||||||||||||||||||||||||||||||||
Db       1202 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1261

Qy        246 s 246
              |
Db       1262 A 1262

RESULT 5
AAT11671
ID   AAT11671 standard; DNA; 1749 BP.
XX
AC   AAT11671;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human plasma carboxypeptidase B.
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   P-PSDB; AAR90293.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;

Alignment Scores:
Pred. No.:              7.05e-136       Length:         1749
Score:                  1272.00         Matches:        246
Percent Similarity:     81.73%          Conservative:   0
Best Local Similarity:  81.73%          Mismatches:     0
Query Match:            95.07%          Indels:         55
DB:                     17              Gaps:           2

US-09-980-881A-4 (1-246) x AAT11671 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502
```

```
Qy        41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy        61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy        81 IleGlyHis---------------------------------------------------- 83
             |||||||||
Db       623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy        83 ---------------------------------------------------------- 83

Db       683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy        84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy       104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy       124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy       144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy       164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy       184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy       204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy       224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
             |||||||||||||||||||||||
Db      1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy       232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||||||||||||||||||||||||||||||||||
Db      1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy       246 s 246
             |
Db      1283 A 1283
```

```
RESULT 6
AAT62846
ID   AAT62846 standard; DNA; 1749 BP.
XX
AC   AAT62846;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B coding sequence.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             41..1312
FT                   /*tag=  a
FT                   /product= Human PCPB
FT   sig_peptide     41..106
FT                   /*tag=  b
FT   mat_peptide     107..1309
FT                   /*tag=  c
FT   misc_binding    134..177
FT                   /*tag=  d
FT                   /bound_moeity= 46_bp_probe
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   P-PSDB; AAW14733.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence encodes human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence 1749 BP; 521 A; 361 C; 342 G; 525 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              7.05e-136       Length:         1749
Score:                  1272.00         Matches:        246
Percent Similarity:     81.73%          Conservative:   0
Best Local Similarity:  81.73%          Mismatches:     0
Query Match:            95.07%          Indels:         55
DB:                     18              Gaps:           2

US-09-980-881A-4 (1-246) x AAT62846 (1-1749)

Qy         1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy        21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy        41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy        61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy        81 IleGlyHis---------------------------------------------------- 83
             |||||||||
```

```
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ------------------------------------------------------------ 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
               |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
               |
Db       1283 A 1283
```

```
RESULT 7
AAQ41001
ID   AAQ41001 standard; cDNA; 1749 BP.
XX
AC   AAQ41001;
XX
DT   24-AUG-1993  (first entry)
XX
DE   Human plasma carboxypeptidase B gene.
XX
KW   PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   sig_peptide     41..106
FT                   /*tag=  a
FT   mat_peptide     107..1312
FT                   /*tag=  b
FT                   /product= PCPB
FT   misc_feature    134..177
FT                   /*tag=  c
FT                   /function= probe
FT                   /note= "used to obtain full-length clones"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   P-PSDB; AAR36273.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Disclosure; Fig 4; 40pp; English.
XX
CC   Human plasma carboxypeptidase B was isolated from human plasma and
CC   partially sequenced. Oligonucleotide primers were designed based on
CC   the partial amino acid sequences. The primers were used in a PCR
CC   amplification to identify cDNA encoding PCPB from a human liver cDNA
CC   library. The PCR product was capable of encoding the first 37 amino
CC   acids of PCPB; a 46mer probe was used to obtain the full-length
CC   sequence which, although disclosed in the specification, is not
CC   claimed. PCPB inhibits the enzymatic conversion by tPA of
CC   plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence 1749 BP; 521 A; 360 C; 343 G; 525 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              2.65e-135      Length:          1749
Score:                  1267.00        Matches:         245
Percent Similarity:     81.40%         Conservative:    0
Best Local Similarity:  81.40%         Mismatches:      1
Query Match:            94.69%         Indels:          55
DB:                     14             Gaps:            2

US-09-980-881A-4 (1-246) x AAQ41001 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ------------------------------------------------------------ 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATGCAGTTCCTCATGC 862
```

```
Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
              |
Db       1283 A 1283
```

RESULT 8
AAF18005
ID AAF18005 standard; DNA; 1400 BP.
XX
AC AAF18005;
XX
DT 14-MAR-2001 (first entry)
XX
DE Lung cancer associated polynucleotide sequence SEQ ID 24.
XX
KW Human; lung cancer associated protein; neuroprotective; cytostatic;
KW cardioactive; immunomodulatory; muscular active; vulnerary;
KW gastrointestinal; nephrotropic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease; ds.
XX
OS Homo sapiens.
XX
PN WO200055180-A2.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05918.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
XX
PI Ruben SM;
XX
DR WPI; 2000-587514/55.
DR P-PSDB; AAB58129.
XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX
PS Claim 1; Page 507; 1425pp; English.
XX
CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective; cytostatic; cardioactive;
CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC general; nephrotropic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the
CC protein or polynucleotide sequences. The lung cancer associated
CC polynucleotide sequences may be used for detection of lung cancer,
CC chromosome identification, as chromosome markers, and for numerous other
CC diagnostic or research purposes. The proteins may be used to treat
CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.
XX
SQ Sequence 1400 BP; 418 A; 281 C; 284 G; 415 T; 2 other;

Alignment Scores:

Pred. No.:	1.8e-93	Length:	1400
Score:	902.00	Matches:	181
Percent Similarity:	76.05%	Conservative:	0
Best Local Similarity:	76.05%	Mismatches:	1
Query Match:	67.41%	Indels:	56
DB:	21	Gaps:	2

US-09-980-881A-4 (1-246) x AAF18005 (1-1400)

```
Qy         65 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 83
              |||||||||||||||   ||||||||||||||||||||||||||||||||||||||||
Db         10 GGAATCCATGCCAGANAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATA 69

Qy         83 ------------------------------------------------------------ 83

Db         70 ACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTAT 129

Qy         84 ----------------------------------------------AsnArgMetTr 87
                                                             ||||||||||
Db        130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACTACTCATGGAAAAAGAATCGAATGTG 189

Qy         87 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAA 249

Qy        107 nPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309

Qy        127 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAG 369

Qy        147 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 AAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTT 429

Qy        167 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 TCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAG 489

Qy        187 rGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 TGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGG 549

Qy        207 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIl 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609

Qy        227 eLysTyrSerPhe---------------------------------------------- 231
              |||||||||||
Db        610 CAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGA 669
```

```
Qy        232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
                |||||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 719

RESULT 9
AAT35760
ID   AAT35760 standard; cDNA; 927 BP.
XX
AC   AAT35760;
XX
DT   09-NOV-1996  (first entry)
XX
DE   Rat mature carboxypeptidase B cDNA.
XX
KW   Carboxypeptidase-B; pro-enzyme; protease; insulin; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   primer_bind     complement (1..50)
FT                   /*tag=  a
FT                   /note= "Mature CPB 5'-end primer"
FT   primer_bind     898..927
FT                   /*tag=  b
FT                   /note= "CPB 3' end primer"
XX
PN   WO9623064-A1.
XX
PD   01-AUG-1996.
XX
PF   25-JAN-1996;   96WO-US00995.
XX
PR   25-JAN-1995;   95US-0378233.
XX
PA   (BIOT-) BIOTECHNOLOGY GEN CORP.
XX
PI   Fulga N,  Gorecki M,  Hartman J,  Mendelovitch S;
XX
DR   WPI; 1996-362688/36.
DR   P-PSDB; AAW00602.
XX
PT   Purified active recombinant carboxypeptidase B prodn. - by
PT   expressing DNA encoding the pro-enzyme, folding and enzymatic
PT   cleavage to give active form, used e.g. for insulin prodn.
XX
PS   Example 1; Page 37-38; 49pp; English.
XX
CC   A cDNA sequence (AAT35760) codes for the mature form (AAW00602) of
CC   rat carboxypeptidase B (CPB).  A full-length sequence including
CC   cDNA (AAT35759) coding for the activation peptide (AAW00601) can
CC   be obtd. by PCR amplification (see also AAT35756 and AAT35758) of cDNA
CC   derived from Sprague-Dawley rat pancreas.  Pro-CPB can be expressed
CC   in host cells, e.g. Escherichia coli, and subsequently recovered,
CC   refolded and cleaved with trypsin to yield the active enzyme.  CPB
CC   produced this way is cheaper than porcine pancreatic enzyme, and is
CC   free of other proteases.
XX
SQ   Sequence 927 BP; 253 A; 226 C; 213 G; 235 T; 0 other;

Alignment Scores:
Pred. No.:              3.72e-60        Length:         927
Score:                  610.00          Matches:        126
Percent Similarity:     60.50%          Conservative:   44
Best Local Similarity:  44.84%          Mismatches:     71
Query Match:            45.59%          Indels:         41
DB:                     17              Gaps:           5

US-09-980-881A-4 (1-246) x AAT35760 (1-927)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||         |||   :::|||::::::         |||   :::|||||||:::
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::||||||::::::|||:::      ||||||::::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                     :::||||||||||||:::   ||||||   :::      |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              :::||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy         81 Ile---------------------------------------------------------- 81
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297

Qy         82 ----------------------------------GlyHis------------------- 83
                                                |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::||||||||||||||   ||||||      |||   :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||   :::      |||||||||   |||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||      ||||||||||||   |||   ||||||:::|||
Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||   |||::::::   ||||||||||:::::::::||||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||||                :::::::::||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
                 :::         |||   :::   :::      |||          |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        224 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 243
              |||   ||||||||||||||||||||         |||      :::||||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        244 Ser 244
              |||
Db        831 TCT 833

RESULT 10
AAQ90601
ID   AAQ90601 standard; DNA; 921 BP.
XX
AC   AAQ90601;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine carboxypeptidase B coding sequence.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB; ds.
XX
OS   Sus scrofa.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             1..921
FT                   /*tag= a
FT                   /product= Porcine carboxypeptidase B.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   P-PSDB; AAR75132.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Claim 1; Page 20; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence 921 BP; 270 A; 238 C; 194 G; 219 T; 0 other;

Alignment Scores:
Pred. No.:              6.12e-57      Length:        921
Score:                  582.00        Matches:       117
Percent Similarity:     59.47%        Conservative:  40
Best Local Similarity:  44.32%        Mismatches:    67
Query Match:            43.50%        Indels:        40
DB:                     16            Gaps:          5

US-09-980-881A-4 (1-246) x AAQ90601 (1-921)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              ||||||:::|||:::::          |||   :::|||   :::   :::|||
Db         16 TATGAGAAGTACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAA 75

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::|||||::::::::::::      ||||||::::::|||            :::|||
Db         76 AATCCAGACCTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATAC 135

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              :::||||||||||   ||||||          |||   ||||||::::::|||||||||
Db        136 CTCCTCAAGGTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGT 192

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
                 ||||||||||||||||||||   |||||||||   ||||||:::
Db        193 TTCCATGCCAGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTT 252

Qy         81 ------------------------------------------------------------ 81

Db        253 CTCACCTATGGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTC 312

Qy         82 --------------------GlyHis------------------AsnArgMetTrpArg 88
                                   |||:::                 ||||||||||||||
Db        313 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 372

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              |||   ||||||      |||          ||||||||||||||   ||||||||||||
Db        373 AAGACCCGCTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT 432

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                      ||||||      |||||||||:::      |||   ||||||||||||
Db        433 ---GATGCTGGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGT 489

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||          |||||||||   |||   ||||||:::|||   |||:::|||   |||
Db        490 GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAAC 549

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
              ::::::   ||||||||||||:::::::::||||||||||||||   |||::::::|||
Db        550 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCT 609

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              ||||||||                 :::::::::   ||||||:::   :::|||
Db        610 TATTCCTATGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCT 669

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
              ||||||:::   :::      |||          |||:::||||||:::|||   |||:::
Db        670 GCCGTGAAAGAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCT 726

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
                 |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db        727 ACAACAATCTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 786

Qy        229 TyrSerPheThr 232
              ||||||||||||
Db        787 TATTCCTTCACC 798

RESULT 11
AAQ90600
ID   AAQ90600 standard; DNA; 1215 BP.
XX
AC   AAQ90600;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine Tyr-His-Met procarboxypeptidase B coding sequence.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB; ds.
XX
OS   Sus scrofa.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1215
FT                   /*tag= a
FT                   /product= Porcine procarboxypeptidase B.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   P-PSDB; AAR75131.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
```

```
PS   Claim 2; Page 20-21; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027). The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding. The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence 1215 BP; 358 A; 300 C; 266 G; 291 T; 0 other;

Alignment Scores:
Pred. No.:              9.21e-57      Length:         1215
Score:                  582.00        Matches:        117
Percent Similarity:     59.47%        Conservative:   40
Best Local Similarity:  44.32%        Mismatches:     67
Query Match:            43.50%        Indels:         40
DB:                     16            Gaps:           5

US-09-980-881A-4 (1-246) x AAQ90600 (1-1215)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              ||||||::|||:::::          |||   ::::||   :::   :::|||
Db        310 TATGAGAAGTACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAA 369

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::||||||:::::::::::::      |||||::::::|||             :::|||
Db        370 AATCCAGACCTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATAC 429

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              :::||||||||||   ||||||          |||   ||||||::::::|||||||||
Db        430 CTCCTCAAGGTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGT 486

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
                 ||||||||||||||||||||   |||||||||   ||||||:::
Db        487 TTCCATGCCAGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTT 546

Qy         81 ------------------------------------------------------------ 81

Db        547 CTCACCTATGGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTC 606

Qy         82 ---------------------GlyHis-----------------AsnArgMetTrpArg 88
                                  |||:::                 ||||||||||||||
Db        607 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 666

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              |||   ||||||      |||          ||||||||||||||   ||||||||||||
Db        667 AAGACCCGCTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT 726

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerCysSerGluThrTyrCys 128
                  ||||||      |||||||||:::      |||   ||||||||||||
Db        727 ---GATGCTGGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGT 783

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||          |||||||||   |||   ||||||::|||   |||:::|||   |||
Db        784 GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAAC 843

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
              ::::::   ||||||||||||:::::::::||||||||||||||   |||::::::|||
Db        844 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCT 903

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              |||||||||          :::::::::   ||||||:::   :::|||
Db        904 TATTCCTATGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCT 963

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
              ||||||:::   :::    |||         |||:::||||||:::|||   |||:::
Db        964 GCCGTGAAAGAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCT 1020

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
              |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db       1021 ACAACAATCTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 1080

Qy        229 TyrSerPheThr 232
              ||||||||||||
Db       1081 TATTCCTTCACC 1092

RESULT 12
AAT42500
ID   AAT42500 standard; DNA; 999 BP.
XX
AC   AAT42500;
XX
DT   17-FEB-1997  (first entry)
XX
DE   Mature HCPB coding sequence with PelB leader sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; non-selective triggering; primer;
KW   polymerase chain reaction; PCR; HP-RNase; Fd; F(ab')2; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..990
FT                   /*tag=  a
FT   sig_peptide     1..66
FT                   /*tag=  b
FT                   /note= "PelB leader sequence"
FT   mat_peptide     67..987
FT                   /*tag=  c
FT                   /note= "mature HCPB"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06174.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 16; Page 137-139; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The mature native HCPB and PelB leader sequence coding sequences
CC   were cloned (using primers AAT42498-99) from pICI1712 (see AAT42497) for
CC   expression in E. coli. The present sequence encodes the mature HCPB and
CC   PelB leader sequence.
XX
SQ   Sequence 999 BP; 263 A; 261 C; 227 G; 248 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              3.66e-55        Length:         999
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
Query Match:            42.38%          Indels:         40
DB:                     17              Gaps:           5

US-09-980-881A-4 (1-246) x AAT42500 (1-999)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::         ||||||:::|||:::           |||   :::|||   :::
Db         67 GCAACTGGTCACTCTTACGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 126

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::|||   ::::::::::::      ||||||::::::||||||
Db        127 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 186

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                     :::|||:::|||||||||   ||||||           |||   ||||||
Db        187 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 243

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        244 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 303

Qy         81 Ile------------------------------------------------------- 81
              :::
Db        304 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 363

Qy         82 ------------------------------------GlyHis------------------ 83
                                                  |||:::
Db        364 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 423

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   ||||||||   ||||||   :::      :::   ||||||||||||||
Db        424 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 483

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
                 ||||||||||||          |||||||||   |||||||||   :::   |||
Db        484 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 540

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
                 ||||||||||||||          |||||||||   |||   ||||||:::|||
Db        541 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 600

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||||:::::::::||||||||||||||
Db        601 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 660

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||               :::::::::   ||||||:::
Db        661 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 720

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
                 :::|||         |||:::   :::      :::         |||:::||||||
Db        721 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 777

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        778 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 837

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::|||||||||||
Db        838 GACCAAGGAATCAGATATTCCTTCACC 864

RESULT 13
AAT42497
ID   AAT42497 standard; DNA; 1053 BP.
XX
AC   AAT42497;
XX
DT   12-FEB-1997  (first entry)
XX
DE   mature HCPB-(His)6-c-myc coding sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1053
FT                   /*tag=  a
FT                   /note= "double stop codon"
FT   sig_peptide     1..66
FT                   /*tag= b
FT                   /note= "PelB leader sequence"
FT   mat_peptide     67..987
FT                   /*tag=  c
FT                   /note= "mature HCPB"
FT   misc_feature    988..1047
FT                   /*tag=  d
FT                   /note= "(His)6-c-myc tag"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06173.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 15; Page 134-136; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The mature native HCPB coding sequence (see AAT42494) was cloned
CC   into the  secretion vector pICI266 (which contains a PelB translation
CC   leader  sequence). To help purification and detection of the expression
CC   construct a C-terminal petpide tag, (His)6-c-myc was added to the enzyme.
CC   The tag  consists of 6 histidines, a tri-peptide linker (EPE) and a
CC   peptide  sequence (EQKLISEEDL) from c-myc which is recognises by antibody
CC   9E10.  The C-terminal is completed by addition of an aspargine. The final
CC    vector, pICI1712, contains the present sequence which encodes the
CC   mature HCPB-(His)6-c-myc protein with PelB leader sequence.
XX
SQ   Sequence 1053 BP; 279 A; 278 C; 241 G; 255 T; 0 other;

Alignment Scores:
Pred. No.:              3.96e-55        Length:         1053
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
```

```
Query Match:          42.38%      Indels:       40
DB:                   17          Gaps:         5

US-09-980-881A-4 (1-246) x AAT42497 (1-1053)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::          ||||||:::|||:::             |||   :::|||   :::
Db         67 GCAACTGGTCACTCTTACGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 126

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::          :::|||  ::::::::::::      ||||||::::::||||||
Db        127 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 186

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                    :::|||:::||||||||||   ||||||           |||   ||||||
Db        187 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 243

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        244 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 303

Qy         81 Ile------------------------------------------------------- 81
              :::
Db        304 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 363

Qy         82 ------------------------------------GlyHis------------------ 83
                                                  |||:::
Db        364 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 423

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   |||||||||   ||||||   :::       :::   ||||||||||||||
Db        424 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 483

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||          |||||||||   |||||||||   :::   |||
Db        484 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 540

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||          |||||||||   |||   ||||||:::|||
Db        541 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 600

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   |||||||||||:::::::::||||||||||||||||
Db        601 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 660

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||||          :::::::::   ||||||:::
Db        661 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 720

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              :::|||          |||:::   :::       :::          |||:::||||||
Db        721 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 777

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   ||||||||   |||
Db        778 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 837

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::||||||||||||
Db        838 GACCAAGGAATCAGATATTCCTTCACC 864

RESULT 14
AAT42494
ID   AAT42494 standard; DNA; 1263 BP.
XX
AC   AAT42494;
XX
DT   12-FEB-1997  (first entry)
XX
DE   Human pancreatic carboxypeptidase B encoding sequence.
XX
KW   ribonuclease; human; bovine; pancreatic; anti-tumour therapy; ADEPT;
KW   mustard-ribonucleotide; antibody directed enzyme prodrug therapy;
KW   anti-neoplastic; prodrug; reverse polarity; ion pair interaction;
KW   reduced immunogenicity; primer; PCR; polymerase chain reaction; HP-RNase;
KW   Fd; F(ab')2; PelB; leader; human carboxypeptidase B; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1248
FT                   /*tag=  a
FT                   /product= preproenzyme
FT   sig_peptide     1..39
FT                   /*tag=  b
FT                   /note= "partial enzyme secretion leader sequence
FT                        (pre-sequence); residues -108 to -96"
FT   sig_peptide     40..324
FT                   /*tag=  c
FT                   /note= "pro-sequence; residues -95 to -1"
FT   mat_peptide     385..1245
FT                   /*tag=  d
FT                   /note= "mature enzyme; encodes residues +1 to +307"
XX
PN   WO9620011-A1.
XX
PD   04-JUL-1996.
XX
PF   21-DEC-1995;   95WO-GB02991.
XX
PR   16-AUG-1995;   95GB-0016810.
PR   23-DEC-1994;   94GB-0026192.
XX
PA   (ZENE ) ZENECA LTD.
XX
PI   Blakey DC,  Boyle FT,  Davies DH,  Eggelte HJ,  Heaton DW;
PI   Hennam JF, Hennequin LFA,  Marsham PR,  Rabin BR,  Slater AM;
PI   Tarragona-Fiol A, Taylorson CJ;
XX
DR   WPI; 1996-321650/32.
DR   P-PSDB; AAW06172.
XX
PT   Two component system for anti-tumour therapy - comprising targeting
PT   moiety linked to mutated enzyme which can transform an
PT   anti:neoplastic prodrug
XX
PS   Reference Example 14; Page 131; 182pp; English.
XX
CC   A two-component system for anti-tumour therapy comprises a targeting
CC   moiety linked to a mutated enzyme which can transform an anti-neoplastic
CC   prodrug. The system is based on antibody directed enzyme prodrug therapy
CC   (ADEPT) using non-naturally occurring mutant forms of host enzymes,
CC   pref. human pancreatic ribonuclease (HP-RNase), (see AAT42478-83).
CC   Alternatively a modified human pancreatic carboxypeptidase B (HCPB) can
CC   be used. The present sequence encodes the native prepro-HCPB (see
CC   AAW06172) contained in vector pICI1698.
XX
SQ   Sequence 1263 BP; 343 A; 306 C; 302 G; 312 T; 0 other;

Alignment Scores:
Pred. No.:              5.18e-55      Length:        1263
Score:                  567.00        Matches:       111
Percent Similarity:     58.36%        Conservative:  46
Best Local Similarity:  41.26%        Mismatches:    72
Query Match:            42.38%        Indels:        40
DB:                     17            Gaps:          5

US-09-980-881A-4 (1-246) x AAT42494 (1-1263)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::          ||||||:::|||:::             |||   :::|||   :::
Db        325 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 384

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::          :::|||  ::::::::::::      ||||||::::::||||||
Db        385 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 444
```

```
Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                  :::|||:::||||||||||   ||||||             |||   ||||||
Db        445 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 501

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        502 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 561

Qy         81 Ile---------------------------------------------------------- 81
              :::
Db        562 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 621

Qy         82 ------------------------------------GlyHis------------------ 83
                                                  |||:::
Db        622 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 681

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   |||||||||   ||||||   :::      :::   ||||||||||||||
Db        682 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 741

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
                 ||||||||||||           |||||||||   |||||||||   :::   |||
Db        742 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 798

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
                 ||||||||||||||         |||||||||   |||   ||||||:::|||
Db        799 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 858

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||||:::::::::||||||||||||||||
Db        859 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 918

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||               :::::::::   ||||||:::
Db        919 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 978

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
                 :::|||          |||:::   :::      :::         |||:::||||||
Db        979 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1035

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db       1036 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1095

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::||||||||||||
Db       1096 GACCAAGGAATCAGATATTCCTTCACC 1122

RESULT 15
AAV41795
ID   AAV41795 standard; DNA; 1263 BP.
XX
AC   AAV41795;
XX
DT   20-NOV-1998  (first entry)
XX
DE   Human pancreatic carboxypeptidase nucleotide sequence.
XX
KW   ss; human; pancreatic carboxypeptidase B; insulin; protein sequencing;
KW   prodrug therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1248
FT                   /*tag=  a
FT                   /product= "Pancreatic carboxypeptidase B"
FT                   /note= "No start codon given"
XX
PN   WO9835988-A1.
XX
PD   20-AUG-1998.
XX
PF   10-FEB-1998;   98WO-GB00415.
XX
PR   29-OCT-1997;   97GB-0022727.
PR   14-FEB-1997;   97GB-0003104.
PR   18-OCT-1997;   97GB-0022003.
XX
PA   (ZENE ) ZENEPA   (ZENE ) ZENECA LTD.
XX
PI   Edge MD;
XX
DR   WPI; 1998-467168/40.
DR   P-PSDB; AAW74476.
XX
PT   New modified pro-domain of carboxy-peptidase B - enhances expression
PT   of co-expressed proteins for production of recombinant
PT   carboxy-peptidase or its fusions with antibodies, used, e.g. in
PT   enzyme prodrug therapy
XX
PS   Example 1; Page 52-53; 83pp; English.
XX
CC   The human pancreatic carboxypeptidase B (CPB) can be co-expressed with a
CC   modified pro-domain of CPB on a separate gene to enhance recombinant
CC   expression.  This process can be used to produce recombinant CPB in
CC   eukaryotic cells, or fusions of CPB with antibody chains.  CPB is used in
CC   insulin production and protein sequencing, while its fusions with
CC   antibodies are useful in antibody-directed enzyme prodrug therapy.  The
CC   modified pro-domain provide increased yields of recombinant CPB, possibly
CC   by protecting the C-terminus against enzymatic degradation or by
CC   increasing intracellular trafficking.
XX
SQ   Sequence 1263 BP; 344 A; 306 C; 301 G; 312 T; 0 other;

Alignment Scores:
Pred. No.:              5.18e-55        Length:         1263
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
Query Match:            42.38%          Indels:         40
DB:                     19              Gaps:           5

US-09-980-881A-4 (1-246) x AAV41795 (1-1263)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::        ||||||:::|||:::          |||   :::|||   :::
Db        325 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 384

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::|||   ::::::::::::      ||||||:::::||||||
Db        385 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 444

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                      :::|||:::||||||||||   ||||||             |||   ||||||
Db        445 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 501

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        502 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 561

Qy         81 Ile---------------------------------------------------------- 81
              :::
Db        562 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 621

Qy         82 ------------------------------------GlyHis------------------ 83
                                                  |||:::
Db        622 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 681

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   |||||||||   ||||||   :::      :::   ||||||||||||||
Db        682 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 741

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
                 ||||||||||||           |||||||||   |||||||||   :::   |||
```

```
Db        742 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 798

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||         |||||||||   |||   ||||||:::|||
Db        799 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGAT 858

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||||:::::::::||||||||||||||||
Db        859 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 918

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              :::::::::||||||||||||                 :::::::::   ||||||:::
Db        919 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 978

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              :::|||         |||:::   :::      :::         |||:::||||||
Db        979 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1035

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db       1036 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1095

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   |||||||:::||||||||||||
Db       1096 GACCAAGGAATCAGATATTCCTTCACC 1122
```

Search completed: January 1, 2004, 20:09:34
Job time : 279.597 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:51:12 ; Search time 112.985 Seconds
(without alignments)
961.011 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135917_26894/app_query.fasta_1.1628
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_115_@runat_31122003_135917_26894 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1289.5	96.4	1625	4	US-09-813-133A-1	Sequence 1, Appli
2	1272	95.1	1272	2	US-08-869-057-1	Sequence 1, Appli
3	1272	95.1	1749	1	US-07-649-591B-2	Sequence 2, Appli
4	1272	95.1	1749	1	US-08-277-540-2	Sequence 2, Appli
5	1272	95.1	1749	1	US-08-430-787A-2	Sequence 2, Appli
6	610	45.6	927	2	US-08-782-760-5	Sequence 5, Appli
7	610	45.6	927	5	PCT-US96-00995-5	Sequence 5, Appli
8	582	43.5	921	1	US-08-696-139-3	Sequence 3, Appli
9	582	43.5	1215	1	US-08-696-139-1	Sequence 1, Appli
10	567	42.4	999	2	US-08-860-882A-67	Sequence 67, Appl
11	567	42.4	999	4	US-09-011-769A-50	Sequence 50, Appl
12	567	42.4	1053	2	US-08-860-882A-64	Sequence 64, Appl
13	567	42.4	1053	4	US-09-011-769A-46	Sequence 46, Appl
14	567	42.4	1263	2	US-08-860-882A-56	Sequence 56, Appl
15	567	42.4	1263	4	US-09-011-769A-38	Sequence 38, Appl
16	567	42.4	1284	2	US-08-860-882A-71	Sequence 71, Appl
17	567	42.4	1284	4	US-09-011-769A-55	Sequence 55, Appl
18	560	41.9	1053	4	US-09-463-451-27	Sequence 27, Appl
c 19	560	41.9	1053	4	US-09-463-451-28	Sequence 28, Appl
20	560	41.9	1059	2	US-08-860-882A-74	Sequence 74, Appl
21	560	41.9	1059	4	US-09-011-769A-59	Sequence 59, Appl
22	559	41.8	1059	2	US-08-860-882A-77	Sequence 77, Appl
23	559	41.8	1059	4	US-09-011-769A-63	Sequence 63, Appl
24	556.5	41.6	1311	4	US-09-675-305-9	Sequence 9, Appli
25	549	41.0	1870	3	US-09-171-945-112	Sequence 112, App
26	549	41.0	2154	3	US-09-171-945-124	Sequence 124, App
27	525.5	39.3	2128	4	US-09-675-305-13	Sequence 13, Appl
28	492	36.8	1200	4	US-09-710-099-7	Sequence 7, Appli
29	471.5	35.2	1050	4	US-09-675-305-11	Sequence 11, Appl
30	465.5	34.8	1257	3	US-08-640-906-1	Sequence 1, Appli
31	465.5	34.8	1257	4	US-09-395-936-1	Sequence 1, Appli
32	465.5	34.8	1311	4	US-09-710-099-5	Sequence 5, Appli
33	459	34.3	1251	3	US-08-640-906-3	Sequence 3, Appli
34	459	34.3	1251	4	US-09-395-936-3	Sequence 3, Appli
35	405.5	30.3	945	4	US-09-710-099-3	Sequence 3, Appli
36	405.5	30.3	945	4	US-09-710-099-11	Sequence 11, Appl
37	403.5	30.2	2247	4	US-09-710-099-15	Sequence 15, Appl
38	379	28.3	1056	4	US-09-710-099-1	Sequence 1, Appli
39	379	28.3	1056	4	US-09-710-099-9	Sequence 9, Appli
40	356	26.6	55827	4	US-09-813-133A-3	Sequence 3, Appli
41	225.5	16.9	741	4	US-09-675-305-5	Sequence 5, Appli
42	198.5	14.8	629	4	US-09-280-116-228	Sequence 228, App
43	169.5	12.7	494	4	US-09-669-751-63	Sequence 63, Appl
c 44	169.5	12.7	515	3	US-08-998-416-125	Sequence 125, App
45	163	12.2	591	3	US-09-331-709-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-813-133A-1
; Sequence 1, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Human
US-09-813-133A-1

Alignment Scores:
Pred. No.: 5.61e-153 Length: 1625
Score: 1289.50 Matches: 244
Percent Similarity: 92.42% Conservative: 0
Best Local Similarity: 92.42% Mismatches: 2
Query Match: 96.38% Indels: 18
DB: 4 Gaps: 1

US-09-980-881A-4 (1-246) x US-09-813-133A-1 (1-1625)

Qy 1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
Db 359 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 418

Qy 21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

```
Db        419 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 478

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        479 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 538

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 598

Qy         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 658

Qy        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GGAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 718

Qy        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 778

Qy        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 838

Qy        161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 898

Qy        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        899 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACC 958

Qy        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1018

Qy        221 TrpIleTyrAspLeuGlyIleLysTyrSerPhe-------------------------- 231
              |||||||||||||||||||||||||||||||
Db       1019 TGGATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACA 1078

Qy        232 ------------------------ThrSerAsnProProValGluLysLeuLeuProLe 243
                                       ||||||||||||||||||||||||||||||||||
Db       1079 TACGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 1138

Qy        243 uSerLeuLys 246
              ||||||||||
Db       1139 GTCTCTAAAA 1148

RESULT 2
US-08-869-057-1
; Sequence 1, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1272 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    PUBLICATION INFORMATION:
;      AUTHORS:  Eaton, Dan L.
;      AUTHORS:  Malloy, Beth E.
;      AUTHORS:  Tsai, Siao P
;      AUTHORS:  Henzel, William
;      AUTHORS:  Drayna, Dennis
;      TITLE:  Isolation, Molecular Cloning, and Partial
;      TITLE:  Characterization of a No. 5985562el Carboxypeptidase B
;      TITLE:  from Human Plasma
;      JOURNAL:  J. Biol. Chem.
;      VOLUME:  266
;      ISSUE:  32
;      PAGES:  21833-21838
;      DATE:  No. 5985562 15-1991
US-08-869-057-1

Alignment Scores:
Pred. No.:                6.17e-151      Length:          1272
Score:                    1272.00        Matches:         246
Percent Similarity:       81.73%         Conservative:    0
Best Local Similarity:    81.73%         Mismatches:      0
Query Match:              95.07%         Indels:          55
DB:                       2              Gaps:            2

US-09-980-881A-4 (1-246) x US-08-869-057-1 (1-1272)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||||
Db        583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 642

Qy         83 ------------------------------------------------------------ 83

Db        643 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        763 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 882

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1062

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1182

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1183 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
              |
Db       1243 A 1243
```

```
RESULT 3
US-07-649-591B-2
; Sequence 2, Application US/07649591B
; Patent No. 5206161
;  GENERAL INFORMATION:
;    APPLICANT:  Dennis Drayna and Daniel Eaton
;    TITLE OF INVENTION:  No. 5206161el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  hybridization probe
;      LOCATION:  133 to 178
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  potential clip site
;      LOCATION:  380 to 382
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  signal sequence
;      LOCATION:  41 to 106
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-07-649-591B-2

Alignment Scores:
Pred. No.:                1.03e-150      Length:          1749
Score:                    1272.00        Matches:         246
Percent Similarity:       81.73%         Conservative:    0
Best Local Similarity:    81.73%         Mismatches:      0
Query Match:              95.07%         Indels:          55
DB:                       1              Gaps:            2

US-09-980-881A-4 (1-246) x US-07-649-591B-2 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||||
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ---------------------------------------------------------- 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042
```

```
Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
              |
Db       1283 A 1283
```

```
RESULT 4
US-08-277-540-2
; Sequence 2, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-277-540-2

Alignment Scores:
Pred. No.:              1.03e-150       Length:         1749
Score:                  1272.00         Matches:        246
Percent Similarity:     81.73%          Conservative:   0
Best Local Similarity:  81.73%          Mismatches:     0
Query Match:            95.07%          Indels:         55
DB:                     1               Gaps:           2

US-09-980-881A-4 (1-246) x US-08-277-540-2 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy         81 IleGlyHis------------------------------------------------------ 83
              ||||||||
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ------------------------------------------------------------ 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
              |
Db       1283 A 1283
```

```
RESULT 5
US-08-430-787A-2
; Sequence 2, Application US/08430787A
```

```
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1749 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-430-787A-2

Alignment Scores:
Pred. No.:                1.03e-150      Length:          1749
Score:                    1272.00        Matches:         246
Percent Similarity:       81.73%         Conservative:    0
Best Local Similarity:    81.73%         Mismatches:      0
Query Match:              95.07%         Indels:          55
DB:                       1              Gaps:            2

US-09-980-881A-4 (1-246) x US-08-430-787A-2 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ---------------------------------------------------------- 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                              ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
              |
Db       1283 A 1283

RESULT 6
US-08-782-760-5
; Sequence 5, Application US/08782760
; Patent No. 5948668
;  GENERAL INFORMATION:
;    APPLICANT:  Hartman, Jacob
;    APPLICANT:  Fulga, Netta
;    APPLICANT:  Mendelovitch, Simona
;    APPLICANT:  Gorecki, Marian
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
```

```
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,760
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/378,233
;      FILING DATE:  25-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
US-08-782-760-5

Alignment Scores:
Pred. No.:              2.05e-67      Length:        927
Score:                  610.00        Matches:       126
Percent Similarity:     60.50%        Conservative:  44
Best Local Similarity:  44.84%        Mismatches:    71
Query Match:            45.59%        Indels:        41
DB:                     2             Gaps:          5

US-09-980-881A-4 (1-246) x US-08-782-760-5 (1-927)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||         |||   :::|||:::::          |||   :::|||||||:::
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              :::         :::|||||:::::|||:::      ||||||:::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                 :::||||||||||||:::   ||||||   :::      |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              :::||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy         81 Ile------------------------------------------------------- 81
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297

Qy         82 ------------------------------------GlyHis------------------ 83
                                                  |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::||||||||||||||   ||||||      |||   :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||   :::      |||||||||   |||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||      ||||||||||   |||   ||||||:::|||
Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||   |||::::::   ||||||||||:::::::::|||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||                :::::::::|||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
                 :::          |||   :::   :::      |||          |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   ||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        224 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 243
              |||   ||||||||||||||||||||||          |||      :::||||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        244 Ser 244
              |||
Db        831 TCT 833

RESULT 7
PCT-US96-00995-5
; Sequence 5, Application PC/TUS9600995
;  GENERAL INFORMATION:
;    APPLICANT:  Bio-Technology General Corp.
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00995
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847-A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  927 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..927
PCT-US96-00995-5

Alignment Scores:
Pred. No.:              2.05e-67      Length:        927
Score:                  610.00        Matches:       126
```

```
Percent Similarity:     60.50%          Conservative:  44
Best Local Similarity:  44.84%          Mismatches:    71
Query Match:            45.59%          Indels:        41
DB:                     5               Gaps:          5

US-09-980-881A-4 (1-246) x PCT-US96-00995-5 (1-927)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||      |||   :::|||::::::      |||   :::|||||:::
Db          1 GCAAGTGGACACAGCTACACCAAGTACAACAACTGGGAAACGATTGAGGCGTGGATTCAA 60

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              :::         :::||||||:::::||:::      ||||||:::::||||||
Db         61 CAAGTTGCCACTGATAATCCAGACCTTGTCACTCAGAGCGTCATTGGAACCACATTTGAA 120

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                    :::||||||||||||:::   ||||||   :::      |||   ||||||
Db        121 GGACGTAACATGTATGTCCTCAAGATT---GGTAAAACTAGACCGAATAAGCCTGCCATC 177

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              :::||||||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        178 TTCATCGATTGTGGTTTCCATGCAAGAGAGTGGATTTCTCCTGCATTCTGTCAGTGGTTT 237

Qy         81 Ile---------------------------------------------------------- 81
              :::
Db        238 GTGAGAGAGGCTGTCCGTACCTATAATCAAGAGATCCACATGAAACAGCTTCTAGATGAA 297

Qy         82 ----------------------------------GlyHis------------------- 83
                                                 |||:::
Db        298 CTGGATTTCTATGTTCTGCCTGTGGTCAACATTGATGGCTATGTCTACACCTGGACTAAG 357

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::||||||||||||||   ||||||      |||   :::   |||:::|||   |||
Db        358 GACAGAATGTGGAGAAAAACCCGCTCTACTATGGCTGGAAGTTCCTGCTTGGGTGTAGAC 417

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
                 ||||||||||   :::      |||||||||   |||||||||   |||   |||
Db        418 CCCAACAGGAATTTT---AATGCTGGCTGGTGTGAAGTGGGAGCTTCTCGGAGTCCCTGC 474

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||      ||||||||||||   |||   ||||||:::|||
Db        475 TCTGAAACTTACTGTGGACCAGCCCCAGAGTCTGAAAAAGAGACAAAGGCCCTGGCAGAT 534

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||   |||::::::   |||||||||||||:::::::::|||||||||||||||
Db        535 TTCATCCGCAACAACCTCTCCACCATCAAGGCCTACCTGACCATCCACTCATACTCACAG 594

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||||          :::::::::||||||||||:::
Db        595 ATGATGCTCTACCCTTACTCCTATGACTACAAACTGCCTGAGAACTATGAGGAATTGAAT 654

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
                 :::         |||   :::   :::      |||         |||:::||||||
Db        655 GCCCTGGTGAAAGGTGCGGCAAAGGAGCTT---GCCACTCTGCATGGCACCAAGTACACA 711

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db        712 TATGGCCCAGGAGCTACAACAATCTATCCTGCTGCTGGGGGATCTGACGACTGGTCTTAT 771

Qy        224 AspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeuLeuProLeu 243
              |||   ||||||||||||||||||||         |||       :::||||||   |||
Db        772 GATCAGGGAATCAAATATTCCTTTACCTTTGAACT-CCGGGATACAGGCTTCTTTGGCTT 830

Qy        244 Ser 244
              |||
Db        831 TCT 833

RESULT 8
US-08-696-139-3
; Sequence 3, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  921 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..921
US-08-696-139-3

Alignment Scores:
Pred. No.:              7.07e-64        Length:        921
Score:                  582.00          Matches:       117
Percent Similarity:     59.47%          Conservative:  40
Best Local Similarity:  44.32%          Mismatches:    67
Query Match:            43.50%          Indels:        40
DB:                     1               Gaps:          5

US-09-980-881A-4 (1-246) x US-08-696-139-3 (1-921)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              ||||||:::|||::::::      |||   :::|||   :::   :::|||
Db         16 TATGAGAAGTACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAA 75

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::||||||:::::::::::      ||||||::::::|||           :::|||
Db         76 AATCCAGACCTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATAC 135

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              :::|||||||||   ||||||            |||   ||||||:::::|||||||||
Db        136 CTCCTCAAGGTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGT 192

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
                 ||||||||||||||||||||   |||||||||   ||||||:::
Db        193 TTCCATGCCAGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTT 252

Qy         81 ---------------------------------------------------------- 81
```

```
Db        253 CTCACCTATGGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTC 312

Qy         82 --------------------GlyHis-----------------AsnArgMetTrpArg 88
                                  |||:::                |||||||||||||||
Db        313 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 372

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
               |||   ||||||      |||         |||||||||||||   ||||||||||||
Db        373 AAGACCCGCTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT 432

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                      ||||||      |||||||||:::      |||   ||||||||||||
Db        433 ---GATGCTGGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGT 489

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
               |||         |||||||||   |||   ||||||:::|||   |||:::|||   |||
Db        490 GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAAC 549

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
               ::::::   |||||||||||:::::::::|||||||||||||||   |||:::::|||
Db        550 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCT 609

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
               |||||||||               :::::::::   ||||||:::   :::|||
Db        610 TATTCCTATGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCT 669

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
               ||||||:::   :::      |||         |||:::|||||:::|||   |||:::
Db        670 GCCGTGAAAGAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCT 726

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
               |||:::|||   |||   ||||||   |||||||||   ||||||   |||||||||
Db        727 ACAACAATCTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 786

Qy        229 TyrSerPheThr 232
               ||||||||||||
Db        787 TATTCCTTCACC 798

RESULT 9
US-08-696-139-1
; Sequence 1, Application US/08696139
; Patent No. 5672496
;  GENERAL INFORMATION:
;    APPLICANT:  Fayerman, Jeffrey T.
;    APPLICANT:  Greenen, David P.
;    APPLICANT:  Hershberger, Charles L.
;    APPLICANT:  Larson, Jeffrey L.
;    APPLICANT:  Sterner, Jane L.
;    APPLICANT:  Zhang, Haichao
;    TITLE OF INVENTION:  DNA SEQUENCES ENCODING PORCINE
;    TITLE OF INVENTION:  PANCREATIC CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  United States of America
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/696,139
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,258
;      FILING DATE:  16-NOV-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gaylo, Paul J.
;      REGISTRATION NUMBER:  36,808
;      REFERENCE/DOCKET NUMBER:  X-8681
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (317) 276-0756
;      TELEFAX:  (317) 276-3861
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1215 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1215
US-08-696-139-1

Alignment Scores:
Pred. No.:                1.11e-63        Length:          1215
Score:                    582.00          Matches:         117
Percent Similarity:       59.47%          Conservative:    40
Best Local Similarity:    44.32%          Mismatches:      67
Query Match:              43.50%          Indels:          40
DB:                       1               Gaps:            5

US-09-980-881A-4 (1-246) x US-08-696-139-1 (1-1215)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              ||||||:::|||:::::          |||   :::|||   :::   :::|||
Db        310 TATGAGAAGTACAACAACTGGGAAACGATCGAGGCTTGGACTAAGCAAGTCACCAGTGAA 369

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::||||||::::::::::::      ||||||:::::|||            :::|||
Db        370 AATCCAGACCTCATCTCTCGCACAGCCATCGGAACTACATTTTTAGGAAACAATATATAC 429

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              :::||||||||||   ||||||            |||   ||||||::::::||||||||
Db        430 CTCCTCAAGGTT---GGCAAACCTGGACCAAATAAGCCTGCCATTTTCATGGACTGTGGT 486

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 81
              ||||||||||||||||||||   |||||||||   ||||||:::
Db        487 TTCCATGCCAGAGAATGGATTTCCCATGCATTTTGCCAGTGGTTTGTGAGAGAGGCTGTT 546

Qy         81 ---------------------------------------------------------- 81

Db        547 CTCACCTATGGATATGAGAGTCACATGACAGAATTCCTCAACAAGCTAGACTTTTATGTC 606

Qy         82 --------------------GlyHis-----------------AsnArgMetTrpArg 88
                                  |||:::                |||||||||||||||
Db        607 TTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAGAACCGAATGTGGAGA 666

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
               |||   ||||||      |||         |||||||||||||   ||||||||||||
Db        667 AAGACCCGCTCTACCAATGCTGGAACTACCTGCATTGGCACAGACCCCAACAGAAATTTT 726

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                      ||||||      |||||||||:::      |||   ||||||||||||
Db        727 ---GATGCTGGGTGGTGCACAACTGGAGCCTCTACAGACCCCTGCGATGAGACTTACTGT 783

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
               |||         |||||||||   |||   ||||||:::|||   |||:::|||   |||
Db        784 GGATCTGCTGCAGAGTCTGAAAAAGAGACCAAGGCCCTGGCTGATTTTATACGCAACAAC 843

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
               ::::::   |||||||||||:::::::::|||||||||||||||   |||:::::|||
Db        844 CTCTCCTCCATCAAAGCATACCTGACGATCCACTCATACTCACAGATGATACTCTACCCT 903

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
               |||||||||               :::::::::   ||||||:::   :::|||
Db        904 TATTCCTATGATTACAAACTCCCCGAGAACAATGCTGAGTTGAATAACCTGGCTAAGGCT 963

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
```

```
Db        964 GCCGTGAAAGAACTT---GCTACACTGTATGGCACCAAGTACACATACGGCCCAGGAGCT 1020

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228

Db       1021 ACAACAATCTATCCTGCTGCTGGGGGCTCTGATGACTGGGCTTATGACCAAGGAATCAAA 1080

Qy        229 TyrSerPheThr 232

Db       1081 TATTCCTTCACC 1092

RESULT 10
US-08-860-882A-67
; Sequence 67, Application US/08860882A
; Patent No. 5985281
; GENERAL INFORMATION:
;   APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;   APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;   APPLICANT:  TARRAGONA-FIOL, ANTONIO
;   APPLICANT:  RABIN, BRIAN ROBERT
;   APPLICANT:  BOYLE, FRANCIS THOMAS
;   APPLICANT:  HENNAM,JOHN FREDERICK
;   APPLICANT:  BLAKELY, DAVID CHARLES
;   APPLICANT:  MARSHAM, PETER ROBERT
;   APPLICANT:  HEATON, DAVID WILLIAM
;   APPLICANT:  DAVIES, DAVID HUW
;   TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;   NUMBER OF SEQUENCES:  77
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;     STREET:  1100 NEW YORK AVENUE, N.W.
;     CITY:  WASHINGTON
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20005
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy Disk
;     COMPUTER:  IBM compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/860,882A
;     FILING DATE:  JUNE 23, 1997
;     CLASSIFICATION:  424
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  DONALD J. BIRD
;     REGISTRATION NUMBER:  25,323
;     REFERENCE/DOCKET NUMBER:  9901/238653
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 861-3027
;     TELEFAX:  (202) 822-0944
;     TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  67:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  999 bases
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
US-08-860-882A-67

Alignment Scores:
Pred. No.:              6.37e-62      Length:        999
Score:                  567.00        Matches:       111
Percent Similarity:     58.36%        Conservative:  46
Best Local Similarity:  41.26%        Mismatches:    72
Query Match:            42.38%        Indels:        40
DB:                     2             Gaps:          5

US-09-980-881A-4 (1-246) x US-08-860-882A-67 (1-999)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

Db         67 GCAACTGGTCACTCTTACGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 126

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

Db        127 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 186

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60

Db        187 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 243

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80

Db        244 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 303

Qy         81 Ile---------------------------------------------------------- 81

Db        304 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 363

Qy         82 ----------------------------------GlyHis------------------- 83

Db        364 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 423

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103

Db        424 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 483

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123

Db        484 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 540

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143

Db        541 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 600

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163

Db        601 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 660

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183

Db        661 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 720

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203

Db        721 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 777

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223

Db        778 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 837

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232

Db        838 GACCAAGGAATCAGATATTCCTTCACC 864

RESULT 11
US-09-011-769A-50
; Sequence 50, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
```

```
;           ZIP: 20005
;       COMPUTER READABLE FORM:
;           MEDIUM TYPE: 1.44 Mb disk
;           COMPUTER: IBM PC compatible
;           OPERATING SYSTEM: PC-DOS/MS-DOS
;           SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;           APPLICATION NUMBER: US/09/011,769A
;           FILING DATE: 13-Feb-1998
;           CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;           APPLICATION NUMBER: PCT/GB96/01975
;           FILING DATE: 13-AUG-1996
;           APPLICATION NUMBER: GB 9612295.7
;           FILING DATE: 12-JUN-1996
;           APPLICATION NUMBER: GB 9611019.2
;           FILING DATE: 25-MAY-1996
;           APPLICATION NUMBER: GB 9516810.0
;           FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 50:
;       SEQUENCE CHARACTERISTICS:
;           LENGTH: 999 base pairs
;           TYPE: nucleic acid
;           STRANDEDNESS: single
;           TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;       FEATURE:
;           NAME/KEY:  CDS
;           LOCATION: 1..987
;       FEATURE:
;           NAME/KEY:  mat_peptide
;           LOCATION: 67..987
;       SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-09-011-769A-50

Alignment Scores:
Pred. No.:              6.37e-62        Length:         999
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
Query Match:            42.38%          Indels:         40
DB:                     4               Gaps:           5

US-09-980-881A-4 (1-246) x US-09-011-769A-50 (1-999)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::          |||||:::|||:::          |||   :::|||   :::
Db         67 GCAACTGGTCACTCTTACGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 126

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::          :::|||  ::::::::::::      |||||:::::|||||
Db        127 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 186

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                    :::|||:::|||||||||   ||||||           |||   ||||||
Db        187 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 243

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              :::::|||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        244 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 303

Qy         81 Ile------------------------------------------------------- 81
              :::
Db        304 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 363

Qy         82 ----------------------------------GlyHis------------------ 83
                                                  |||:::
Db        364 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 423

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   |||||||||   ||||||   :::      :::   ||||||||||||||
Db        424 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 483

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||          |||||||||   |||||||||   :::   |||
Db        484 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 540

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||          |||||||||   |||   ||||||:::|||
Db        541 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 600

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||:::::::::||||||||||||||||
Db        601 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 660

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                  :::::::::||||||||||||               :::::::::   ||||||:::
Db        661 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 720

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              :::|||          |||:::   :::      :::          |||:::||||||
Db        721 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 777

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   ||||||||   |||
Db        778 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 837

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::||||||||||||
Db        838 GACCAAGGAATCAGATATTCCTTCACC 864

RESULT 12
US-08-860-882A-64
; Sequence 64, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  64:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1053 bases
;      TYPE:  nucleic acid
```

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-64

Alignment Scores:
Pred. No.:              6.93e-62        Length:         1053
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
Query Match:            42.38%          Indels:         40
DB:                     2               Gaps:           5

US-09-980-881A-4 (1-246) x US-08-860-882A-64 (1-1053)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::         ||||||:::|||:::          |||   :::|||   :::
Db         67 GCAACTGGTCACTCTTACGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 126

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::|||   ::::::::::::      |||||:::::|||||
Db        127 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 186

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                    :::|||:::||||||||||   ||||||          |||   ||||||
Db        187 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 243

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        244 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 303

Qy         81 Ile---------------------------------------------------------- 81
              :::
Db        304 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 363

Qy         82 ----------------------------------GlyHis------------------ 83
                                                  |||:::
Db        364 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 423

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   ||||||||   ||||||   :::      :::   ||||||||||||||
Db        424 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 483

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||          |||||||||   |||||||||   :::   |||
Db        484 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 540

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||          |||||||||   |||  ||||||:::|||
Db        541 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 600

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||:::::::::||||||||||||||
Db        601 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 660

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              :::::::::||||||||||||          :::::::::   ||||||:::
Db        661 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 720

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              :::|||         |||:::  :::      :::         |||:::||||||
Db        721 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 777

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   ||||||||   |||
Db        778 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 837

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::||||||||||||
Db        838 GACCAAGGAATCAGATATTCCTTCACC 864

RESULT 13
US-09-011-769A-46
; Sequence 46, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 46:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1053 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;       FEATURE:
;            NAME/KEY:  CDS
;            LOCATION: 1..1047
;       FEATURE:
;            NAME/KEY:  mat_peptide
;            LOCATION: 67..1047
;       SEQUENCE DESCRIPTION: SEQ ID NO: 46:
US-09-011-769A-46

Alignment Scores:
Pred. No.:              6.93e-62        Length:         1053
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
Query Match:            42.38%          Indels:         40
DB:                     4               Gaps:           5

US-09-980-881A-4 (1-246) x US-09-011-769A-46 (1-1053)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::         ||||||:::|||:::          |||   :::|||   :::
Db         67 GCAACTGGTCACTCTTACGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 126

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::|||   ::::::::::::      |||||:::::|||||
Db        127 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 186

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                    :::|||:::||||||||||   ||||||          |||   ||||||
Db        187 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 243
```

```
Qy          61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
               ::::::|||||||||   ||||||||||||||||||||||||||||||||   ||||||
Db         244 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 303

Qy          81 Ile---------------------------------------------------------- 81
               :::
Db         304 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 363

Qy          82 ----------------------------------GlyHis------------------- 83
                                                   |||:::
Db         364 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 423

Qy          84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
               :::|||   ||||||||   ||||||   :::      :::   ||||||||||||||
Db         424 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 483

Qy         104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
               ||||||||||||          |||||||||   |||||||||   :::   |||
Db         484 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 540

Qy         124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
               ||||||||||||||||          |||||||||   |||   ||||||:::|||
Db         541 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 600

Qy         144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
               |||:::|||      ::::::   ||||||||||:::::::::||||||||||||||||
Db         601 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 660

Qy         164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
               :::::::::|||||||||||          :::::::::   ||||||:::
Db         661 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 720

Qy         184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
               :::|||         |||:::   :::      :::          |||:::||||||
Db         721 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 777

Qy         204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
               :::|||   |||:::   |||:::|||   |||   ||||||   |||||||||   |||
Db         778 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 837

Qy         224 AspLeuGlyIleLysTyrSerPheThr 232
               |||   ||||||:::|||||||||||
Db         838 GACCAAGGAATCAGATATTCCTTCACC 864

RESULT 14
US-08-860-882A-56
; Sequence 56, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  56:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1263 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-860-882A-56

Alignment Scores:
Pred. No.:                9.31e-62        Length:          1263
Score:                    567.00          Matches:         111
Percent Similarity:       58.36%          Conservative:    46
Best Local Similarity:    41.26%          Mismatches:      72
Query Match:              42.38%          Indels:          40
DB:                       2               Gaps:            5

US-09-980-881A-4 (1-246) x US-08-860-882A-56 (1-1263)

Qy           1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
               |||:::      ||||||:::|||:::          |||  :::|||   :::
Db         325 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 384

Qy          21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
               :::         :::|||   ::::::::::::      ||||||:::::|||||
Db         385 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 444

Qy          41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                  :::|||:::|||||||||   ||||||           |||   ||||||
Db         445 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 501

Qy          61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
               ::::::|||||||||   ||||||||||||||||||||||||||||||||   ||||||
Db         502 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 561

Qy          81 Ile---------------------------------------------------------- 81
               :::
Db         562 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 621

Qy          82 ----------------------------------GlyHis------------------- 83
                                                   |||:::
Db         622 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 681

Qy          84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
               :::|||   ||||||||   ||||||   :::      :::   ||||||||||||||
Db         682 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 741

Qy         104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
               ||||||||||||          |||||||||   |||||||||   :::   |||
Db         742 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 798

Qy         124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
               ||||||||||||||||          |||||||||   |||   ||||||:::|||
Db         799 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 858

Qy         144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
               |||:::|||      ::::::   ||||||||||:::::::::||||||||||||||||
Db         859 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 918

Qy         164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
               :::::::::|||||||||||          :::::::::   ||||||:::
```

```
Db        919 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 978

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203

Db        979 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1035

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223

Db       1036 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1095

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232

Db       1096 GACCAAGGAATCAGATATTCCTTCACC 1122

RESULT 15
US-09-011-769A-38
; Sequence 38, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 38:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 1263 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: other nucleic acid
;       SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-011-769A-38

Alignment Scores:
Pred. No.:              9.31e-62        Length:         1263
Score:                  567.00          Matches:        111
Percent Similarity:     58.36%          Conservative:   46
Best Local Similarity:  41.26%          Mismatches:     72
Query Match:            42.38%          Indels:         40
DB:                     4               Gaps:           5

US-09-980-881A-4 (1-246) x US-09-011-769A-38 (1-1263)
```

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

Db        325 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 384

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

Db        385 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 444

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60

Db        445 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 501

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80

Db        502 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 561

Qy         81 Ile------------------------------------------------------------ 81

Db        562 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 621

Qy         82 ----------------------------------GlyHis----------------------- 83

Db        622 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 681

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103

Db        682 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 741

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123

Db        742 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 798

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143

Db        799 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAAACCAAGGCCCTGGCTGAT 858

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163

Db        859 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 918

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183

Db        919 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 978

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203

Db        979 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1035

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223

Db       1036 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1095

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232

Db       1096 GACCAAGGAATCAGATATTCCTTCACC 1122

Search completed: January  2, 2004, 09:26:07
Job time : 125.985 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 20:10:07 ; Search time 598.309 Seconds
(without alignments)
1423.134 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2263443 seqs, 1730637950 residues

Total number of hits satisfying chosen parameters: 4526886

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135919_26949/app_query.fasta_1.1628
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09980881_@CGN_1_1_704_@runat_31122003_135919_26949
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1272	95.1	1728	10	US-09-880-107-2396	Sequence 2396, Ap
2	1177	88.0	1272	13	US-10-379-836-1	Sequence 1, Appli
3	902	67.4	1400	9	US-09-925-302-24	Sequence 24, Appl
4	556.5	41.6	1311	14	US-10-200-344-9	Sequence 9, Appli
5	556.5	41.6	1993	12	US-10-274-639-33	Sequence 33, Appl
6	552	41.3	1254	15	US-10-229-546-3	Sequence 3, Appli
7	552	41.3	1622	12	US-10-429-802-21	Sequence 21, Appl
8	552	41.3	1622	15	US-10-229-546-1	Sequence 1, Appli
9	552	41.3	1633	13	US-10-341-434-187	Sequence 187, App
10	550.5	41.1	1125	9	US-09-888-615-2	Sequence 2, Appli
11	550.5	41.1	1332	16	US-10-176-306-75	Sequence 75, Appl
12	550.5	41.1	1603	16	US-10-176-306-73	Sequence 73, Appl
13	549	41.0	1870	10	US-09-910-059-112	Sequence 112, App
14	549	41.0	2154	10	US-09-910-059-124	Sequence 124, App
15	547.5	40.9	1332	10	US-09-954-456-1141	Sequence 1141, Ap
16	543.5	40.6	1826	13	US-10-252-157-453	Sequence 453, App
17	525.5	39.3	2128	14	US-10-200-344-13	Sequence 13, Appl
18	492	36.8	1200	15	US-10-200-910-7	Sequence 7, Appli
19	471.5	35.2	1050	14	US-10-200-344-11	Sequence 11, Appl
20	465.5	34.8	1254	9	US-09-925-297-58	Sequence 58, Appl
21	465.5	34.8	1260	13	US-10-345-680-66	Sequence 66, Appl
22	465.5	34.8	1311	15	US-10-200-910-5	Sequence 5, Appli
23	465.5	34.8	1380	10	US-09-969-347-209	Sequence 209, App
24	465.5	34.8	1380	13	US-10-345-680-64	Sequence 64, Appl
25	458	34.2	2786	11	US-09-946-374-233	Sequence 233, App
26	458	34.2	2786	13	US-10-015-387A-233	Sequence 233, App
27	458	34.2	2786	13	US-10-006-130A-233	Sequence 233, App
28	458	34.2	2786	13	US-10-199-672-307	Sequence 307, App
29	458	34.2	2786	13	US-10-006-172A-233	Sequence 233, App
30	458	34.2	2786	13	US-10-187-749-307	Sequence 307, App
31	458	34.2	2786	13	US-10-194-457-307	Sequence 307, App
32	458	34.2	2786	13	US-10-184-642-307	Sequence 307, App
33	458	34.2	2786	13	US-10-196-747-307	Sequence 307, App
34	458	34.2	2786	13	US-10-015-392A-233	Sequence 233, App
35	458	34.2	2786	13	US-10-017-253A-233	Sequence 233, App
36	458	34.2	2786	13	US-10-173-689-307	Sequence 307, App
37	458	34.2	2786	13	US-10-173-690-307	Sequence 307, App
38	458	34.2	2786	13	US-10-173-691-307	Sequence 307, App
39	458	34.2	2786	13	US-10-173-692-307	Sequence 307, App
40	458	34.2	2786	13	US-10-173-694-307	Sequence 307, App
41	458	34.2	2786	13	US-10-173-698-307	Sequence 307, App
42	458	34.2	2786	13	US-10-173-699-307	Sequence 307, App
43	458	34.2	2786	13	US-10-173-707-307	Sequence 307, App
44	458	34.2	2786	13	US-10-174-569-307	Sequence 307, App
45	458	34.2	2786	13	US-10-174-583-307	Sequence 307, App

ALIGNMENTS

RESULT 1
US-09-880-107-2396
; Sequence 2396, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2396
; LENGTH: 1728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

```
;  OTHER INFORMATION: Genbank Accession No. US20020142981A1 M75106
US-09-880-107-2396

Alignment Scores:
Pred. No.:                3.06e-160      Length:          1728
Score:                    1272.00        Matches:         246
Percent Similarity:       81.73%         Conservative:    0
Best Local Similarity:    81.73%         Mismatches:      0
Query Match:              95.07%         Indels:          55
DB:                       10             Gaps:            2

US-09-980-881A-4 (1-246) x US-09-880-107-2396 (1-1728)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 421

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 481

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 541

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 601

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        602 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 661

Qy         83 ---------------------------------------------------------- 83

Db        662 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 721

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 781

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 841

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 901

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 961

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1021

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1141

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1142 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1201

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||||||||||||||||||||||||||||||||||
Db       1202 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1261

Qy        246 s 246
              |
Db       1262 A 1262

RESULT 2
US-10-379-836-1
; Sequence 1, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
;   LENGTH: 1272
;   TYPE: DNA
;   ORGANISM: Papio hamadryas
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1269)
US-10-379-836-1

Alignment Scores:
Pred. No.:                1.13e-147      Length:          1272
Score:                    1177.00        Matches:         228
Percent Similarity:       78.41%         Conservative:    8
Best Local Similarity:    75.75%         Mismatches:      10
Query Match:              87.97%         Indels:          55
DB:                       13             Gaps:            2

US-09-980-881A-4 (1-246) x US-10-379-836-1 (1-1272)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||::::::|||||||||||||||||||||||||||||||||||:::|||
Db        403 CTTATAACTGAGAAGTATCCTGATATGCTTACAAAAATCCACATTGGATCCTCCTATGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||:::
Db        463 AAGCACCCACTTTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATG 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCCCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        583 ATAGGCCATATAACTGAATACTACGGGATAATAGGGGAATATACCAATCTTCTGAGGCAT 642

Qy         83 ---------------------------------------------------------- 83

Db        643 GTGGATTTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCGTTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||    ||||||
Db        763 CTGAACAGGAACTTTGCGTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTTCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||    ||||||||||:::
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGCGAAGGCGGTGGCTAAT 882
```

```
Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCACATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATCGTGTTTCCATATTCCTATACTCGAAGCAAAAGCAAAGACCACGAGGAATTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||:::|||||||||||||||   |||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTCAGAAAACCAGTAAAAATATCAGGTATACA 1062

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||   ||||||||||||||||||||||||||||||   ||||||||||||||||
Db       1063 CATGGCCGTGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGCGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCAAATACGGATTC 1182

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||   |||:::  ||||||||||||||||||||
Db       1183 TTGCTGCCTGAGCGTTACATCAAACCCACTTGTAAAGACGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
              |
Db       1243 A 1243
```

RESULT 3
US-09-925-302-24
; Sequence 24, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA104
; CURRENT APPLICATION NUMBER: US/09/925,302
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05918
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 896
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (25)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-302-24

Alignment Scores:

Pred. No.:	1.36e-110	Length:	1400
Score:	902.00	Matches:	181
Percent Similarity:	76.05%	Conservative:	0
Best Local Similarity:	76.05%	Mismatches:	1
Query Match:	67.41%	Indels:	56
DB:	9	Gaps:	2

US-09-980-881A-4 (1-246) x US-09-925-302-24 (1-1400)

```
Qy         65 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis--- 83
              ||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db         10 GGAATCCATGCCAGANAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGCCATATA 69

Qy         83 ------------------------------------------------------------ 83

Db         70 ACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGATTTCTAT 129

Qy         84 ----------------------------------------------AsnArgMetTr 87
                                                            ||||||||||
Db        130 GTTATGCCGGTGGTTAATGTGGATGGTTATGAACTACTCATGGAAAAAGAATCGAATGTG 189

Qy         87 pArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAs 107
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        190 GAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAATAGGAA 249

Qy        107 nPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTy 127
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        250 CTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAAACCTA 309

Qy        127 rCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgAr 147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        310 CTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTGAGAAG 369

Qy        147 gAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPh 167
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        370 AAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATAGTGTT 429

Qy        167 eProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSe 187
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        430 TCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTAGCCAG 489

Qy        187 rGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGl 207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        490 TGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGCCATGG 549

Qy        207 ySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIl 227
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        550 CTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTGGGCAT 609

Qy        227 eLysTyrSerPhe------------------------------------------- 231
              |||||||||||
Db        610 CAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTGCCGGA 669

Qy        232 -----ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
                   ||||||||||||||||||||||||||||||||||||||||||||||
Db        670 GCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 719
```

RESULT 4
US-10-200-344-9
; Sequence 9, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Nehls, Michael
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0047-USA
; CURRENT APPLICATION NUMBER: US/10/200,344
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: US/09/675,305
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1311
; TYPE: DNA
; ORGANISM: homo sapiens
US-10-200-344-9

Alignment Scores:

Pred. No.:	3.86e-64	Length:	1311
Score:	556.50	Matches:	113
Percent Similarity:	53.45%	Conservative:	42
Best Local Similarity:	38.97%	Mismatches:	77
Query Match:	41.59%	Indels:	58
DB:	14	Gaps:	5

```
US-09-980-881A-4 (1-246) x US-10-200-344-9 (1-1311)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              ||||||  ||||||||||||  ||||||  :::|||:::      :::   :::
Db        406 TATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAAACT 465

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              |||      ::::::            ||||||   |||:::|||         |||:::
Db        466 CACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTTTTT 525

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              :::||||||:::   |||:::           |||   |||:::|||||||||||||||
Db        526 ATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGTGGT 582

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              ||||||||||||||||||  ||||||||||||  ||||||:::
Db        583 ATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCTCTT 642

Qy         81 ---------------------------------------------------------- 81

Db        643 CTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTATATC 702

Qy         82 ----------------------------------------GlyHisAsnArgMetTrpArg 88
                                                      ::::::|||   ||||||
Db        703 ATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGGAGA 762

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              |||   ||||||      :::          |||   |||   |||   ||||||||:::
Db        763 AAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAACTGG 822

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                  |||   ||||||:::||||||||||||        |||   :::||||||||
Db        823 AAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATACTGT 879

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||   :::||||||||||||||||||||||||||||||:::|||||||||::::::
Db        880 GGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAACAC 939

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
                  |||:::||||||:::|||   |||:::|||:::|||   ::::::::::|||
Db        940 AGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTATCCC 999

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              |||||||||      :::      :::      :::      |||  :::
Db       1000 TATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTATAAA 1059

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
              ||||||   |||::::::                ||||||  .:::|||      |||
Db       1060 GCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCCTCC 1116

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
                 |||||||||::::::   |||          ||||||   |||      ||||||
Db       1117 ACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATACCT 1176

Qy        229 TyrSerPhe---------------------------------------------------- 231
              |||:::|||
Db       1177 TATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAGATG 1236

Qy        232 -ThrSerAsnProProValGluLysLeu 240
               :::|||||||||||||||:::||||||
Db       1237 CTCATCAAACCCACCTGTACAGAAACTA 1264

RESULT 5
US-10-274-639-33
; Sequence 33, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 33
;   LENGTH: 1993
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US200302323349A1 4948403CB1
US-10-274-639-33

Alignment Scores:
Pred. No.:                7.68e-64        Length:          1993
Score:                    556.50          Matches:         113
Percent Similarity:       53.45%          Conservative:    42
Best Local Similarity:    38.97%          Mismatches:      77
Query Match:              41.59%          Indels:          58
DB:                       12              Gaps:            5

US-09-980-881A-4 (1-246) x US-10-274-639-33 (1-1993)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              ||||||  ||||||||||||  ||||||  :::|||:::      :::   :::
Db        743 TATGAAGTTTATCACTCCTTAGAAGAAATTCAAAATTGGATGCATCATCTGAATAAAACT 802

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              |||      ::::::            ||||||   |||:::|||         |||:::
Db        803 CACTCAGGCCTCATTCACATGTTCTCTATTGGAAGATCATATGAGGGAAGATCTCTTTTT 862

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              :::||||||:::   |||:::           |||   |||:::|||||||||||||||
Db        863 ATTTTAAAGCTG---GGCAGACGATCACGACTCAAAAGAGCTGTTTGGATAGACTGTGGT 919

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              ||||||||||||||||||  ||||||||||||  ||||||:::
Db        920 ATTCATGCAAGAGAATGGATTGGTCCTGCCTTTTGTCAGTGGTTTGTAAAAGAAGCTCTT 979

Qy         81 ---------------------------------------------------------- 81

Db        980 CTAACATATAAGAGTGACCCAGCCATGAGAAAAATGTTGAATCATCTATATTTCTATATC 1039

Qy         82 ----------------------------------------GlyHisAsnArgMetTrpArg 88
```

```
Db       1040 ATGCCTGTGTTTAACGTCGATGGATACCATTTTAGTTGGACCAATGATCGATTTTGGAGA 1099

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108

Db       1100 AAAACAAGGTCAAGGAACTCAAGGTTTCGCTGCCGTGGAGTGGATGCCAATAGAAACTGG 1159

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128

Db       1160 AAAGTGAAG---TGGTGTGATGAAGGAGCTTCTATGCACCCTTGTGATGACACATACTGT 1216

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148

Db       1217 GGCCCTTTTCCAGAATCTGAGCCGGAAGTGAAGGCTGTAGCTAACTTCCTTCGAAAACAC 1276

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168

Db       1277 AGAAAGCACATTAGGGCTTATCTCTCCTTTCATGCATATGCTCAGATGTTACTGTATCCC 1336

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188

Db       1337 TATTCTTACAAATATGCAACAATTCCCAATTTTAGATGTGTGGAATCTGCAGCTTATAAA 1396

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208

Db       1397 GCTGTGAATGCACTTCAGTCAGTA---TACGGGGTACGATACAGATATGGACCAGCCTCC 1453

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228

Db       1454 ACAACGTTGTATGTGAGCTCTGGTAGCTCAATGGATTGGGCCTACAAAAATGGAATACCT 1513

Qy        229 TyrSerPhe------------------------------------------------- 231

Db       1514 TATGCATTTGCTTTCGAACTACGTGACACTGGATATTTTGGATTTTTACTCCCAGAGATG 1573

Qy        232 -ThrSerAsnProProValGluLysLeu 240

Db       1574 CTCATCAAACCCACCTGTACAGAAACTA 1601
```

RESULT 6
US-10-229-546-3
; Sequence 3, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
; APPLICANT: Weich, Nadine S.
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Millennium Pharmaceuticals, Inc.
; TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
; TITLE OF INVENTION: FAMILY MEMBER AND USES THEREFOR
; FILE REFERENCE: MPI01-156P1RNM
; CURRENT APPLICATION NUMBER: US/10/229,546
; CURRENT FILING DATE: 2002-08-28
; PRIOR APPLICATION NUMBER: 60/316,575
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1254)
US-10-229-546-3

Alignment Scores:

Pred. No.:	1.45e-63	Length:	1254
Score:	552.00	Matches:	109
Percent Similarity:	58.56%	Conservative:	45
Best Local Similarity:	41.44%	Mismatches:	69
Query Match:	41.26%	Indels:	40
DB:	15	Gaps:	5

US-09-980-881A-4 (1-246) x US-10-229-546-3 (1-1254)

```
Qy           6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25

Db         346 TACGCAAAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAG 405

Qy          26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45

Db         406 TATCCTGAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATAT 465

Qy          46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65

Db         466 GTTCTGAAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGC 522

Qy          66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81

Db         523 ATTCACGCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACC 582

Qy          81 ------------------------------------------------------------ 81

Db         583 AAAACTTATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATT 642

Qy          82 --------------------GlyHis-----------------AsnArgMetTrpArg 88

Db         643 CTTCCTGTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGA 702

Qy          89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108

Db         703 AAAAATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT 762

Qy         109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128

Db         763 ---AATGCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGG 819

Qy         129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148

Db         820 GGCTCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCAC 879

Qy         149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168

Db         880 CTGAATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCC 939

Qy         169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188

Db         940 TATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATT 999

Qy         189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208

Db        1000 GGCACTGATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAA 1056

Qy         209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228

Db        1057 TCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAA 1116

Qy         229 TyrSerPhe 231

Db        1117 CACACATTT 1125
```

RESULT 7
US-10-429-802-21
; Sequence 21, Application US/10429802
; Publication No. US20030228285A1
; GENERAL INFORMATION:
; APPLICANT: HUNG, MIEN-CHIE
; APPLICANT: WONG, KA YIN
; APPLICANT: ZOU, YIYU
; TITLE OF INVENTION: BIPARTITE T-CELL FACTOR (TCF)-RESPONSIVE PROMOTER
; FILE REFERENCE: UTSC:752US
; CURRENT APPLICATION NUMBER: US/10/429,802
; CURRENT FILING DATE: 2003-05-05
; PRIOR APPLICATION NUMBER: 60/377,672
; PRIOR FILING DATE: 2002-05-03
; NUMBER OF SEQ ID NOS: 54

```
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
;   LENGTH: 1622
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-429-802-21

Alignment Scores:
Pred. No.:              2.21e-63        Length:         1622
Score:                  552.00          Matches:        109
Percent Similarity:     58.56%          Conservative:   45
Best Local Similarity:  41.44%          Mismatches:     69
Query Match:            41.26%          Indels:         40
DB:                     12              Gaps:           5

US-09-980-881A-4 (1-246) x US-10-429-802-21 (1-1622)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              |||   :::|||::::::      :::|||   :::|||   |||   :::   ::::::
Db        346 TACGCAAAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAG 405

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::|||:::|||:::::::::|||   ||||||||::   |||      |||||||||
Db        406 TATCCTGAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATAT 465

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              |||||||||:::   |||::::::          :::   |||||:::::|||||||||
Db        466 GTTCTGAAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGC 522

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              ||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        523 ATTCACGCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACC 582

Qy         81 ------------------------------------------------------------ 81

Db        583 AAAACTTATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATT 642

Qy         82 ---------------------GlyHis-----------------AsnArgMetTrpArg 88
                                   |||:::                 ||||||||||||||
Db        643 CTTCCTGTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGA 702

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              ||||||||||||         |||:::  ||||||||||||||||||||||||||||||
Db        703 AAAAATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT 762

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                 :::   |||                ::::::      |||::::::   |||
Db        763 ---AATGCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGG 819

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||      ||||||||||||   |||   |||||||||   :::|||:::|||   :::
Db        820 GGCTCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCAC 879

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
              :::|||:::||||||  ||||||:::  ||||||||||||||   ::::::||||||
Db        880 CTGAATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCC 939

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              |||   ||||||                ::::||||||:::|||:::   ||||||
Db        940 TATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATT 999

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
                       :::   |||             |||||||   :::|||
Db       1000 GGCACTGATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAA 1056

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
                 |||:::|||         |||         ||||||  ||||||||||||||||||
Db       1057 TCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAA 1116

Qy        229 TyrSerPhe 231
              :::::|||
Db       1117 CACACATTT 1125
```

```
RESULT 8
US-10-229-546-1
; Sequence 1, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1622
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1254)
US-10-229-546-1

Alignment Scores:
Pred. No.:              2.21e-63        Length:         1622
Score:                  552.00          Matches:        109
Percent Similarity:     58.56%          Conservative:   45
Best Local Similarity:  41.44%          Mismatches:     69
Query Match:            41.26%          Indels:         40
DB:                     15              Gaps:           5

US-09-980-881A-4 (1-246) x US-10-229-546-1 (1-1622)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              |||   :::|||::::::      :::|||   :::|||   |||   :::   ::::::
Db        346 TACGCAAAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAG 405

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::|||:::|||:::::::::|||   ||||||||::   |||      |||||||||
Db        406 TATCCTGAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATAT 465

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              |||||||||:::   |||::::::          :::   |||||:::::|||||||||
Db        466 GTTCTGAAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGC 522

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              ||||||||||||||||||:::|||||||||||||||   ||||||:::
Db        523 ATTCACGCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACC 582

Qy         81 ------------------------------------------------------------ 81

Db        583 AAAACTTATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATT 642

Qy         82 ---------------------GlyHis-----------------AsnArgMetTrpArg 88
                                   |||:::                 ||||||||||||||
Db        643 CTTCCTGTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGA 702

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              ||||||||||||         |||:::  ||||||||||||||||||||||||||||||
Db        703 AAAAATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT 762

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                 :::   |||                ::::::      |||::::::   |||
Db        763 ---AATGCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGG 819

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||      ||||||||||||   |||   |||||||||   :::|||:::|||   :::
Db        820 GGCTCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCAC 879

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
```

```
              :::|||:::||||||   ||||||:::   ||||||||||||||   ::::::||||||
Db        880 CTGAATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCC 939

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              |||   ||||||              :::||||||:::|||:::   ||||||
Db        940 TATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATT 999

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
                       :::     |||         |||||||||   :::|||
Db       1000 GGCACTGATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAA 1056

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
                 |||:::|||          |||          ||||||   ||||||||||||||||||
Db       1057 TCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAA 1116

Qy        229 TyrSerPhe 231
              :::::|||
Db       1117 CACACATTT 1125

RESULT 9
US-10-341-434-187
; Sequence 187, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 187
;   LENGTH: 1633
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (12)..(1262)
;   OTHER INFORMATION:
US-10-341-434-187

Alignment Scores:
Pred. No.:              2.23e-63        Length:         1633
Score:                  552.00          Matches:        109
Percent Similarity:     58.56%          Conservative:   45
Best Local Similarity:  41.44%          Mismatches:     69
Query Match:            41.26%          Indels:         40
DB:                     13              Gaps:           5

US-09-980-881A-4 (1-246) x US-10-341-434-187 (1-1633)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              |||   :::|||::::::      :::|||   :::|||   |||   :::   ::::::
Db        357 TACGCAAAATACAATAATTGGGAAAAGATTGTGGCTTGGACTGAAAAGATGATGGATAAG 416

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::|||:::|||:::::::::|||   ||||||||:::   |||      |||||||||
Db        417 TATCCTGAAATGGTCTCTCGTATTAAAATTGGATCTACTGTTGAAGATAATCCACTATAT 476

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
              |||||||||:::   |||::::::         :::   ||||||::::::||||||||||
Db        477 GTTCTGAAGATT---GGGGAAAAGAATGAAAGAAGAAAGGCTATTTTTATGGATTGTGGC 533

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              |||||||||||||||||:::|||||||||||||||   ||||||:::
Db        534 ATTCACGCACGAGAATGGGTCTCCCCAGCATTCTGCCAGTGGTTTGTCTATCAGGCAACC 593

Qy         81 ------------------------------------------------------------ 81
```

```
Db        594 AAAACTTATGGGAGAAACAAAATTATGACCAAACTCTTGGACCGAATGAATTTTTACATT 653

Qy         82 ---------------------GlyHis-----------------AsnArgMetTrpArg 88
                                   |||:::                 ||||||||||||||
Db        654 CTTCCTGTGTTCAATGTTGATGGATATATTTGGTCATGGACAAAGAACCGCATGTGGAGA 713

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              ||||||||||||          |||:::   ||||||||||||||||||||||||||||
Db        714 AAAAATCGTTCCAAGAACCAAAACTCCAAATGCATCGGCACTGACCTCAACAGGAATTTT 773

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                 :::     |||          ::::::      |||::::::   |||
Db        774 ---AATGCTTCATGGAACTCCATTCCTAACACCAATGACCCATGTGCAGATAACTATCGG 830

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||       ||||||||||||   |||   |||||||||   :::|||:::|||   :::
Db        831 GGCTCTGCACCAGAGTCCGAGAAAGAGACGAAAGCTGTCACTAATTTCATTAGAAGCCAC 890

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
              :::|||:::||||||   ||||||:::   ||||||||||||||   ::::::||||||
Db        891 CTGAATGAAATCAAGGTTTACATCACCTTCCATTCCTACTCCCAGATGCTATTGTTTCCC 950

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              |||   ||||||              :::||||||:::|||:::   ||||||
Db        951 TATGGATATACATCAAAACTGCCACCTAACCATGAGGACTTGGCCAAAGTTGCAAAGATT 1010

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
                       :::     |||         |||||||||   :::|||
Db       1011 GGCACTGATGTTCTA---TCAACTCGATATGAAACCCGCTACATCTATGGCCCAATAGAA 1067

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
                 |||:::|||          |||          ||||||   ||||||||||||||||||
Db       1068 TCAACAATTTACCCGATATCAGGTTCTTCTTTAGACTGGGCTTATGACCTGGGCATCAAA 1127

Qy        229 TyrSerPhe 231
              :::::|||
Db       1128 CACACATTT 1136

RESULT 10
US-09-888-615-2
; Sequence 2, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 1125
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-888-615-2

Alignment Scores:
Pred. No.:              1.93e-63        Length:         1125
Score:                  550.50          Matches:        104
Percent Similarity:     56.44%          Conservative:   45
Best Local Similarity:  39.39%          Mismatches:     76
Query Match:            41.14%          Indels:         39
DB:                     9               Gaps:           4

US-09-980-881A-4 (1-246) x US-09-888-615-2 (1-1125)
```

```
Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              |||      ||||||   :::   |||||||||   |||:::      |||:::|||:::
Db        139 TATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAG 198

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::   :::::::::|||:::      :::|||   :::::::|||   :::|||:::|||
Db        199 TACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATGTAT 258

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
                 ||||||:::|||                |||      ||||||:::|||||||||
Db        259 TATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGA 318

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              ||||||||||||||||||:::||||||||||||   ||||||:::
Db        319 ATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTA 378

Qy         81 ------------------------------------------------------------ 81

Db        379 CAAAACCATAAAGACAACTCAAGTATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTC 438

Qy         82 ---------------------GlyHis-----------------AsnArgMetTrpArg 88
                                   |||:::                 :::|||:::||||||
Db        439 CTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGG 498

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
              |||:::||||||   :::   |||      ||:|  ||||||||||||||||||||||
Db        499 AAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC 558

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
                 :::      ||||||      ||||||||   :::        :::|||:::|||
Db        559 ---AATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGT 615

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
              |||      |||   ||||||||||||   |||||||||||||||||:::
Db        616 GGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAG 675

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
                 :::   |||      ::::::::::||||||||||||   |||   |||:::   |||
Db        676 AAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCT 735

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
              |||   ||||||::::::||||||   :::|||   |||:::      |||      :::
Db        736 TACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAG 795

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
              |||      |||:::   |||          |||   |||      |||      :::
Db        796 GCAGCAAATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCA 852

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
              :::   ||||||   :::   |||          ||||||      |||:::||||||
Db        853 GATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC 912

Qy        229 TyrSerPheThr 232
              :::|||:::|||
Db        913 TTCTCATATACG 924
```

```
RESULT 11
US-10-176-306-75
; Sequence 75, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 75
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-176-306-75
```

Alignment Scores:

Pred. No.:	2.55e-63	Length:	1332
Score:	550.50	Matches:	104
Percent Similarity:	56.44%	Conservative:	45
Best Local Similarity:	39.39%	Mismatches:	76
Query Match:	41.14%	Indels:	39
DB:	16	Gaps:	4

US-09-980-881A-4 (1-246) x US-10-176-306-75 (1-1332)

```
Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
              |||      ||||||   :::   |||||||||   |||:::      |||:::|||:::
Db        529 TATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAG 588

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
              :::   :::::::::|||:::      :::|||   :::::::|||   :::|||:::|||
Db        589 TACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATATAT 648

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
                 ||||||:::|||                |||      ||||||:::|||||||||
Db        649 TATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGA 708

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle------------ 81
              ||||||||||||||||||:::||||||||||||   ||||||:::
Db        709 ATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTA 768

Qy         81 ------------------------------------------------------------ 81

Db        769 CAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTC 828

Qy         82 ---------------------GlyHis-----------------AsnArgMetTrpArg 88
                                   |||:::                 :::|||:::||||||
Db        829 CTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGG 888
```

```
Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db        889 AAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC 948

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
Db        949 ---AATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGT 1005

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
Db       1006 GGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAG 1065

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db       1066 AAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCT 1125

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
Db       1126 TACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAG 1185

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
Db       1186 GCAGCAAATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCA 1242

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228
Db       1243 GATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC 1302

Qy        229 TyrSerPheThr 232
Db       1303 TTCTCATATACG 1314

RESULT 12
US-10-176-306-73
; Sequence 73, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glucksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 73
;   LENGTH: 1603
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (272)...(1603)
US-10-176-306-73

Alignment Scores:
Pred. No.:                3.45e-63        Length:          1603
Score:                    550.50          Matches:         104
Percent Similarity:       56.44%          Conservative:    45
Best Local Similarity:    39.39%          Mismatches:      76
Query Match:              41.14%          Indels:          39
DB:                       16              Gaps:            4

US-09-980-881A-4 (1-246) x US-10-176-306-73 (1-1603)

Qy          6 TyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGluPheIleThrGluArg 25
Db        800 TATAACATATACCACCCCATGGGAGAGATCTATGAGTGGATGAGAGAGATCAGTGAGAAG 859

Qy         26 HisProAspMetLeuThrLysIleHisIleGlySerSerPheGluLysTyrProLeuTyr 45
Db        860 TACAAGGAAGTGGTGACACAGCATTTCCTAGGAGTGACCTATGAGACCCACCCCATATAT 919

Qy         46 ValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIleAspCysGly 65
Db        920 TATCTGAAGATCAGCCAACCATCTGGTAATCCCAAGAAAATCATTTGGATGGACTGTGGA 979

Qy         66 IleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIle---------- 81
Db        980 ATTCACGCCAGAGAATGGATTGCTCCTGCTTTTTGCCAATGGTTCGTCAAAGAAATTCTA 1039

Qy         81 ---------------------------------------------------------- 81
Db       1040 CAAAACCATAAAGACAACTCAAGGATACGCAAGCTCCTTAGGAACCTGGACTTCTATGTC 1099

Qy         82 --------------------GlyHis------------------AsnArgMetTrpArg 88
Db       1100 CTTCCAGTTCTTAACATAGATGGTTATATCTACACTTGGACAACTGATCGTCTTTGGAGG 1159

Qy         89 LysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPhe 108
Db       1160 AAATCCCGTTCACCCCATAATAATGGCACATGTTTTGGGACGGATCTCAATCGAAATTTC 1219

Qy        109 AlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCys 128
Db       1220 ---AATGCATCTTGGTGTAGTATTGGTGCCTCTAGAAACTGCCAAGATCAAACATTCTGT 1276

Qy        129 GlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsn 148
Db       1277 GGGACAGGGCCAGTGTCTGAACCAGAGACTAAAGCTGTTGCCAGCTTCATAGAGAGCAAG 1336

Qy        149 IleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPhePro 168
Db       1337 AAGGATGATATTTTGTGCTTCCTGACCATGCACTCTTATGGGCAGTTAATTCTCACACCT 1396

Qy        169 TyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGlu 188
Db       1397 TACGGCTACACCAAAAATAAATCAAGTAACCACCCAGAAATGATTCAAGTTGGACAGAAG 1456

Qy        189 AlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySer 208
```

```
Db       1457 GCAGCAAATGCATTG---AAAGCAAAGTATGGAACCAATTATAGAGTTGGATCGAGTGCA 1513

Qy        209 GluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLys 228

Db       1514 GATATTTTATATGCCTCATCAGGGTCTTCAAGAGATTGGGCCCGAGACATTGGGATTCCC 1573

Qy        229 TyrSerPheThr 232

Db       1574 TTCTCATATACG 1585

RESULT 13
US-09-910-059-112
; Sequence 112, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 112
;   LENGTH: 1870
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: expected PCR insert
US-09-910-059-112

Alignment Scores:
Pred. No.:              7.07e-63        Length:         1870
Score:                  549.00          Matches:        108
Percent Similarity:     57.62%          Conservative:   47
Best Local Similarity:  40.15%          Mismatches:     74
Query Match:            41.03%          Indels:         40
DB:                     10              Gaps:           5

US-09-980-881A-4 (1-246) x US-09-910-059-112 (1-1870)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20

Db        934 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 993

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40

Db        994 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 1053

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60

Db       1054 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 1110

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80

Db       1111 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 1170

Qy         81 Ile---------------------------------------------------------- 81

Db       1171 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 1230

Qy         82 ----------------------------------GlyHis------------------- 83
```

```
Db       1231 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 1290

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103

Db       1291 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 1350

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123

Db       1351 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 1407

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143

Db       1408 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 1467

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163

Db       1468 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 1527

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183

Db       1528 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 1587

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203

Db       1588 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1644

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223

Db       1645 TATGGCCCGGGAGCTACAACAATCTATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTAT 1704

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232

Db       1705 GACCAAGGAATCAGATATTCCTTCACC 1731

RESULT 14
US-09-910-059-124
; Sequence 124, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody, a
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 124
;   LENGTH: 2154
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-124

Alignment Scores:
Pred. No.:              8.91e-63        Length:         2154
Score:                  549.00          Matches:        108
Percent Similarity:     57.62%          Conservative:   47
Best Local Similarity:  40.15%          Mismatches:     74
Query Match:            41.03%          Indels:         40
DB:                     10              Gaps:           5
```

```
US-09-980-881A-4 (1-246) x US-09-910-059-124 (1-2154)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::         |||||||:::|||:::             |||   :::|||   :::
Db        331 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 390

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::|||   ::::::::::::      ||||||:::::::||||||
Db        391 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 450

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                     :::|||:::|||||||||   ||||||             |||   ||||||
Db        451 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 507

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::|||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        508 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 567

Qy         81 Ile------------------------------------------------------------ 81
              :::
Db        568 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 627

Qy         82 -----------------------------------GlyHis------------------ 83
                                                  |||:::
Db        628 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 687

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   |||||||||   ||||||   :::       :::   ||||||||||||||
Db        688 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCTAGCTGCATTGGCACAGAC 747

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||            |||||||||   |||||||||   :::   |||
Db        748 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 804

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||          |||||||||   |||   ||||||:::|||
Db        805 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 864

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||||:::::::::||||||||||||||||
Db        865 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 924

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 :::::::::|||||||||||||                 :::::::::   ||||||:::
Db        925 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 984

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              :::|||         |||:::   :::       :::             |||:::||||||
Db        985 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1041

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   :::   |||         ||||||   |||
Db       1042 TATGGCCCGGGAGCTACAACAATCTATCCTTCTGCTGGGACTTCTAAAGACTGGGCTTAT 1101

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::|||||||||||||
Db       1102 GACCAAGGAATCAGATATTCCTTCACC 1128

RESULT 15
US-09-954-456-1141
; Sequence 1141, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1141
;   LENGTH: 1332
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-1141

Alignment Scores:
Pred. No.:              6.45e-63        Length:         1332
Score:                  547.50          Matches:        110
Percent Similarity:     57.99%          Conservative:   46
Best Local Similarity:  40.89%          Mismatches:     72
Query Match:            40.92%          Indels:         41
DB:                     10              Gaps:           6

US-09-980-881A-4 (1-246) x US-09-954-456-1141 (1-1332)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||:::         |||||||:::|||:::             |||   :::|||   :::
Db        354 GCAACAGGACACAGTTATGAGAAGTACAACAAGTGGGAAACGATAGAGGCTTGGACTCAA 413

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::         :::|||   ::::::::::::      ||||||:::::::||||||
Db        414 CAAGTCGCCACTGAGAATCCAGCCCTCATCTCTCGCAGTGTTATCGGAACCACATTTGAG 473

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
                     :::|||:::|||||||||   ||||||             |||   ||||||
Db        474 GGACGCGCTATTTACCTCCTGAAGGTT---GGCAAAGCTGGACAAAATAAGCCTGCCATT 530

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ::::::|||||||||   ||||||||||||||||||||||||||||||   ||||||
Db        531 TTCATGGACTGTGGTTTCCATGCCAGAGAGTGGATTTCTCCTGCATTCTGCCAGTGGTTT 590

Qy         81 Ile------------------------------------------------------------ 81
              :::
Db        591 GTAAGAGAGGCTGTTCGTACCTATGGACGTGAGATCCAAGTGACAGAGCTTCTCGACAAG 650

Qy         82 -----------------------------------GlyHis------------------ 83
                                                  |||:::
Db        651 TTAGACTTTTATGTCCTGCCTGTGCTCAATATTGATGGCTACATCTACACCTGGACCAAG 710

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              :::|||   |||||||||   ||||||   :::       :::   ||||||||||||||
Db        711 AGCCGATTTTGGAGAAAGACTCGCTCCACCCATACTGGATCT---AGCATTGGCACAGAC 767

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||            |||||||||   |||||||||   :::   |||
Db        768 CCCAACAGAAATTTT---GATGCTGGTTGGTGTGAAATTGGAGCCTCTCGAAACCCCTGT 824

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||          |||||||||   |||   ||||||:::|||
Db        825 GATGAAACTTACTGTGGACCTGCCGCAGAGTCTGAAAAGGAGACCAAGGCCCTGGCTGAT 884

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||:::|||      ::::::   ||||||||||||:::::::::||||||||||||||||
```

```
Db        885 TTCATCCGCAACAAACTCTCTTCCATCAAGGCATATCTGACAATCCACTCGTACTCCCAA 944

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              :::::::::|||||||||||||                :::::::::   ||||||:::
Db        945 ATGATGATCTACCCTTACTCATATGCTTACAAACTCGGTGAGAACAATGCTGAGTTGAAT 1004

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              :::|||          |||:::   :::      :::          |||:::||||||
Db       1005 GCCCTGGCTAAAGCTACTGTGAAAGAACTT---GCCTCACTGCACGGCACCAAGTACACA 1061

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              :::|||   |||:::   |||:::|||   |||   ||||||   ||||||||   |||
Db       1062 TATGGCCCGGGAGCTACAACAATCTATCCTGCTGCTGGGGGCTCTGACGACTGGGCTTAT 1121

Qy        224 AspLeuGlyIleLysTyrSerPheThr 232
              |||   ||||||:::|||||||||||
Db       1122 GACCAAGGAATCAGATATTCCTTCACC 1148
```

Search completed: January 2, 2004, 10:05:26
Job time : 613.309 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:47:17 ; Search time 2179.08 Seconds
(without alignments)
2743.777 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135916_26873/app_query.fasta_1.1628
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_5167_@runat_31122003_135916_26873 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1112	83.1	1446	11	AK004045	AK004045 Mus muscu
2	978	73.1	852	10	BG569281	BG569281 602588573
3	831.5	62.1	688	9	AV651709	AV651709 AV651709
4	828.5	61.9	806	10	BG567264	BG567264 602589745
5	827.5	61.8	577	14	CB154355	CB154355 K-EST0212
c 6	812.5	60.7	973	9	AL536068	AL536068 AL536068
7	795	59.4	510	14	CB156851	CB156851 K-EST0215
8	793.5	59.3	971	9	AL536069	AL536069 AL536069
9	732	54.7	434	14	CB164014	CB164014 K-EST0225
c 10	698.5	52.2	723	9	AW413090	AW413090 uq52c03.x
c 11	685.5	51.2	772	9	AI255929	AI255929 ui83d11.x
12	676.5	50.6	692	10	BF384322	BF384322 602046670
c 13	659.5	49.3	747	9	AI255397	AI255397 ui94a04.x
14	656.5	49.1	713	12	BI246210	BI246210 602958724
15	652.5	48.8	505	14	CB161965	CB161965 K-EST0222
16	649	48.5	645	10	BG642020	BG642020 pg11c.pk0
17	631	47.2	546	14	H61865	H61865 yr02b04.r1
c 18	630	47.1	756	9	AI595707	AI595707 uj68d08.x
19	624	46.6	747	10	BG565100	BG565100 602583720
20	616.5	46.1	932	13	BX451532	BX451532 BX451532
21	607	45.4	378	9	AV651204	AV651204 AV651204
22	607	45.4	1041	11	AK008678	AK008678 Mus muscu
23	606	45.3	853	12	BI144644	BI144644 602909932
24	589	44.0	846	13	BU488930	BU488930 603972157
25	581.5	43.5	927	13	BU487280	BU487280 603972169
26	580.5	43.4	1816	11	AK078883	AK078883 Mus muscu
27	578	43.2	750	13	BU487283	BU487283 604126713
28	577.5	43.2	1429	11	AK003061	AK003061 Mus muscu
29	565	42.2	736	13	BU488223	BU488223 604127073
30	563.5	42.1	1004	14	BY704293	BY704293 BY704293
31	563	42.1	714	13	BU488385	BU488385 604124863
c 32	561.5	42.0	843	10	BG216526	BG216526 RST36106
33	561	41.9	1176	9	AB030423	AB030423 AB030423
34	561	41.9	1453	11	AK088906	AK088906 Mus muscu
35	560.5	41.9	466	9	AA244760	AA244760 mx05d08.r
36	560	41.9	897	14	CB202958	CB202958 AGENCOURT
37	555.5	41.5	984	14	CB207599	CB207599 AGENCOURT
38	553	41.3	766	13	BU490586	BU490586 604131409
39	552	41.3	820	14	CA776082	CA776082 io99d12.y
40	550	41.1	871	13	BU416594	BU416594 603670981
c 41	549	41.0	864	14	CB755833	CB755833 AGENCOURT
42	548	41.0	809	13	BU488219	BU488219 604126116
43	547.5	40.9	805	13	BU487175	BU487175 604126902
44	547	40.9	616	12	BI064948	BI064948 pgf1n.pk0
45	547	40.9	797	13	BU489339	BU489339 604131182

ALIGNMENTS

RESULT 1
AK004045
LOCUS AK004045 1446 bp mRNA linear HTC 05-DEC-2002
DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1110032P04 product:carboxypeptidase B2 (plasma), full insert sequence.
ACCESSION AK004045
VERSION AK004045.1 GI:12835067
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S. and Hayashizaki,Y.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409 (6821), 685-690 (2001)
MEDLINE 21085660
PUBMED 11217851
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1446)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTVN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence[5' GAGAGAGAGAAGGATCCAAGAGCTCAATTAATTTAATTAAACCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.
FEATURES Location/Qualifiers
source 1. .1446
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:1110032P04"
/db_xref="MGI:1900711"
/db_xref="taxon:10090"
/clone="1110032P04"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
CDS 42. .1310
/note="unnamed protein product; carboxypeptidase B2 (plasma) (MGD|MGI:1891837, GB|NM_019775, evidence: BLASTN, 99%, match=1419)
putative"
/codon_start=1
/protein_id="BAB23141.1"
/db_xref="GI:12835068"
/db_xref="MGI:1891837"
/translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT 421 a 329 c 311 g 385 t
ORIGIN

Alignment Scores:

Pred. No.:	6.91e-116	Length:	1446
Score:	1112.00	Matches:	217
Percent Similarity:	75.75%	Conservative:	11
Best Local Similarity:	72.09%	Mismatches:	18
Query Match:	83.11%	Indels:	55
DB:	11	Gaps:	2

US-09-980-881A-4 (1-246) x AK004045 (1-1446)

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        381 GCCTCCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAA 440

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 |||||||||:::||||||||||||||||    ||||||:::||||||||||||||||||
Db        441 GTCATAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAG 500

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        501 AAGTACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATC 560
```

```
Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        561 TGGATCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTC 620

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||:::
Db        621 ATAGGCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCAC 680

Qy         83 ------------------------------------------------------------ 83

Db        681 GTGGATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAG 740

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              |||||||||||||||||||||||||||   :::   ||||||   |||:::||||||||||
Db        741 AATCGAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAC 800

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        801 CTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 860

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        861 TCTGAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 920

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||:::   ||||||||||||||||||||||||||||||||||||||
Db        921 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAA 980

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db        981 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCT 1040

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||      :::||||||||||||||||||
Db       1041 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACA 1100

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||   |||||||||:::||||||||||||||||||||   ||||||||||||||||
Db       1101 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 1160

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1161 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 1220

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||   |||:::||||||   |||   ||||||||
Db       1221 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 1280

Qy        246 s 246
              |
Db       1281 A 1281
```

```
RESULT 2
BG569281
LOCUS       BG569281               852 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602588573F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4722354 5',
            mRNA sequence.
ACCESSION   BG569281
VERSION     BG569281.1  GI:13576934
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1580  row: b  column: 19
            High quality sequence stop: 751.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4722354"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
                     /note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and
                     3' adaptors were used in cloning as follows: 5' adaptor
                     sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size 1.85
                     kb (range 1.0-4.0 kb). 15/15 colonies contained  inserts
                     by PCR. This library was enriched for full-length clones
                     and was constructed by Clontech  Laboratories (Palo Alto,
                     CA). Note: this is a NIH_MGC Library."
BASE COUNT      251 a    177 c    187 g    237 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                6.15e-101      Length:          852
Score:                    978.00         Matches:         197
Percent Similarity:       76.15%         Conservative:    1
Best Local Similarity:    75.77%         Mismatches:      5
Query Match:              73.09%         Indels:          58
DB:                       10             Gaps:            2

US-09-980-881A-4 (1-246) x BG569281 (1-852)

Qy         43 ProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIleTrpIle 62
              ||||||||||||||||||||||||||||||||||||||||   ||||||||||||||||||||
Db          2 CCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATATGGATT 61

Qy         63 AspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGly 82
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 GACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTCATAGGC 121

Qy         83 His---------------------------------------------------------- 83
              |||
Db        122 CATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTTGTGGAT 181

Qy         84 ------------------------------------------------------AsnArg 85
                                                                    ||||||
Db        182 TTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAGAATCGA 241

Qy         86 MetTrp-ArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAs 105
              |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 ATGTGGACGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAA 301

Qy        105 nArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGl 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 TAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGA 361

Qy        125 uThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLe 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 AACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTT 421

Qy        145 uArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIl 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 GAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATAT 481

Qy        165 eValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVa 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGT 541
```

```
Qy        185 lAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGl 205
              ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db        542 AGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACATGC 601

Qy        205 yHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr-AspL 225
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db        602 GCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGCATT 661

Qy        225 euGlyIleLysTyrSerPhe---------------------------------------- 231
              |||||||||:::|||||||||
Db        662 TGGGCATCAGATATTCGTTTACAATTGAACCTTCGAGATACGGGCACATAACGGATTCTT 721

Qy        232 --------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeu 245
                           ||||||   ||||||||||||||||||||||||||||||||
Db        722 GCTGCCGGAGCGTTACATCAAA-CCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTA 776
```

RESULT 3
AV651709

```
LOCUS       AV651709                688 bp    mRNA    linear   EST 15-JAN-2002
DEFINITION  AV651709 GLC Homo sapiens cDNA clone GLCCSF10 3', mRNA sequence.
ACCESSION   AV651709
VERSION     AV651709.1  GI:9872723
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 688)
  AUTHORS   Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
            Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
            Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
            Hu,G., Gu,J., Chen,Z. and Han,Z.
  TITLE     Insight into hepatocellular carcinogenesis at transcriptome level
            by comparing gene expression profiles of hepatocellular carcinoma
            with those of corresponding noncancerous liver
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
  MEDLINE   21625106
   PUBMED   11752456
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn
            This clone is available at CHGC in Shanghai.
FEATURES             Location/Qualifiers
     source          1. .688
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GLCCSF10"
                     /tissue_type="corresponding non cancerous liver tissue"
                     /dev_stage="Adult"
                     /lab_host="SOLR"
                     /clone_lib="GLC"
                     /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      208 a    145 c    151 g    184 t
ORIGIN
```

Alignment Scores:

Pred. No.:	2.4e-84	Length:	688
Score:	831.50	Matches:	162
Percent Similarity:	89.50%	Conservative:	0
Best Local Similarity:	89.50%	Mismatches:	1
Query Match:	62.14%	Indels:	18
DB:	9	Gaps:	1

US-09-980-881A-4 (1-246) x AV651709 (1-688)

```
Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 105

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 165

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 225

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 285

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 345

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 405

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 465

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db        466 GATTTGGGCATCAAATATTCGTTTACAATTGAACCTCGAGATACCGGCACATACCGATTC 525

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||||||||||||||||||||   ||||||||||
Db        526 TTTGTGGCGGAGCGTTACATCAAACCCACCTGTTGAGAAGCTTTTGGCGCTGTCTCTTAA 585

Qy        246 s 246
              |
Db        586 A 586
```

RESULT 4
BG567264

```
LOCUS       BG567264                806 bp    mRNA    linear   EST 10-APR-2001
DEFINITION  602589745F1 NIH_MGC_76 Homo sapiens cDNA clone IMAGE:4723690 5',
            mRNA sequence.
ACCESSION   BG567264
VERSION     BG567264.1  GI:13574917
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 806)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1583  row: j  column: 11
            High quality sequence stop: 772.
FEATURES             Location/Qualifiers
     source          1. .806
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4723690"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NIH_MGC_76"
```

/note="Organ: liver; Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.85 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

BASE COUNT 239 a 158 c 162 g 247 t
ORIGIN

Alignment Scores:

Pred. No.:	6.78e-84	Length:	806
Score:	828.50	Matches:	161
Percent Similarity:	89.94%	Conservative:	0
Best Local Similarity:	89.94%	Mismatches:	0
Query Match:	61.92%	Indels:	18
DB:	10	Gaps:	1

US-09-980-881A-4 (1-246) x BG567264 (1-806)

```
Qy         86 MetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsn 105
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          2 ATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTGAAT 61

Qy        106 ArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGlu 125
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 AGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCGGAA 121

Qy        126 ThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPheLeu 145
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 ACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTCTTG 181

Qy        146 ArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIle 165
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        182 AGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCATATA 241

Qy        166 ValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuVal 185
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 GTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTAGTA 301

Qy        186 AlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGly 205
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        302 GCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACACATGGC 361

Qy        206 HisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeu 225
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 CATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGATTTG 421

Qy        226 GlyIleLysTyrSerPhe---------------------------------------- 231
              |||||||||||||||||
Db        422 GGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTGCTG 481

Qy        232 ----------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
                       ||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 CCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 536
```

RESULT 5
CB154355
LOCUS CB154355 577 bp mRNA linear EST 29-JAN-2003
DEFINITION K-EST0212246 L17N670205 Homo sapiens cDNA clone L17N670205-10-E10 5', mRNA sequence.
ACCESSION CB154355
VERSION CB154355.1 GI:28139354
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 577)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 10 row: E column: 10
High quality sequence stop: 577.
FEATURES Location/Qualifiers
source 1. .577
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="L17N670205-10-E10"
/sex="F"
/lab_host="Top10F'"
/clone_lib="L17N670205"
/note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI; Site_2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."
BASE COUNT 175 a 122 c 128 g 152 t
ORIGIN

Alignment Scores:

Pred. No.:	5.23e-84	Length:	577
Score:	827.50	Matches:	161
Percent Similarity:	89.44%	Conservative:	0
Best Local Similarity:	89.44%	Mismatches:	1
Query Match:	61.85%	Indels:	18
DB:	14	Gaps:	1

US-09-980-881A-4 (1-246) x CB154355 (1-577)

```
Qy         85 ArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 60

Qy        105 AsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSer 124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 120

Qy        125 GluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPhe 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 180

Qy        145 LeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHis 164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 240

Qy        165 IleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeu 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 300

Qy        185 ValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHis 204
              ||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db        301 GTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACAT 360

Qy        205 GlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAsp 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 420

Qy        225 LeuGlyIleLysTyrSerPhe---------------------------------------- 231
              |||||||||||||||||
Db        421 TTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTCTTG 480
```

```
Qy        232 ------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
                           ||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAAA 538

RESULT 6
AL536068/c
LOCUS       AL536068                973 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL536068 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE21 3-PRIME, mRNA sequence.
ACCESSION   AL536068
VERSION     AL536068.2  GI:31260938
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 973)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12799561.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6944.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF022AC11NP1&cluster=6944.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF022AC11NP1.
FEATURES              Location/Qualifiers
     source           1. .973
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DF022YE21"
                      /tissue_type="FETAL BRAIN"
                      /dev_stage="fetal"
                      /clone_lib="Homo sapiens FETAL BRAIN"
                      /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                      was primed with a NotI-oligo(dT) primer. Five prime end
                      enriched, double-strand cDNA was digested with Not I and
                      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                      vector. Library was not normalized."
BASE COUNT      284 a    193 c    188 g    250 t     58 others
ORIGIN

Alignment Scores:
Pred. No.:              6.13e-82      Length:         973
Score:                  812.50        Matches:        159
Percent Similarity:     88.46%        Conservative:   2
Best Local Similarity:  87.36%        Mismatches:     2
Query Match:            60.72%        Indels:         19
DB:                     9             Gaps:           1

US-09-980-881A-4 (1-246) x AL536068 (1-973)

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        954 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATTGGAACAGAC 895

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        894 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 835

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        834 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 775

Qy        144 PheLeuArgArgAsnIle-AsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGl 163
              |||||||||||||||||| :::|||||||||||||||||||||||||||||||||||||
Db        774 TTCTTGAGAAGAAATATCAAGCCAGATTAAAGCATACATCAGCATGCATTCATACTCCCA 715

Qy        163 nHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSe 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        714 GCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTC 655

Qy        183 rLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrTh 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        654 TCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATAC 595

Qy        203 rHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTy 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        594 ACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTA 535

Qy        223 rAspLeuGlyIleLysTyrSerPheThr------------------------------ 232
              |||||||||||||   :::|||||||||
Db        534 TGATTTGGGCATCAATTTTTCGTTTACAATTGAACTTCGAGTTACGGGCACATACGGATT 475

Qy        233 --------------------SerAsnProProValGluLysLeuLeuProLeuSerLeuL 246
                                  |||   ||||||||||||||||||||||||||||||||
Db        474 CTTGCTGCCGGGGCGTTTCCTCAACCCCCCCTGTAGAGAAGCTTTTGCCGCTGTCTCTAA 415

Qy        246 ys 246
              ||
Db        414 AA 413

RESULT 7
CB156851
LOCUS       CB156851                510 bp    mRNA    linear   EST 29-JAN-2003
DEFINITION  K-EST0215830 L17N670205n1 Homo sapiens cDNA clone
            L17N670205n1-4-H06 5', mRNA sequence.
ACCESSION   CB156851
VERSION     CB156851.1  GI:28141979
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 510)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 4  row: H  column: 06
            High quality sequence stop: 510.
FEATURES              Location/Qualifiers
     source           1. .510
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="L17N670205n1-4-H06"
                      /sex="F"
                      /lab_host="Top10F'"
                      /clone_lib="L17N670205n1"
                      /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                      Site_2: NotI; The library was contributed by the Soares
                      laboratory and it was constructed as described by Bonaldo,
                      M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                      6(9): 791-806. RNA was prepared from harvested cell
                      culture."
BASE COUNT      158 a    109 c    113 g    130 t
ORIGIN

Alignment Scores:
```

```
Pred. No.:              2.23e-80        Length:         510
Score:                  795.00          Matches:        147
Percent Similarity:     99.32%          Conservative:   0
Best Local Similarity:  99.32%          Mismatches:     1
Query Match:            59.42%          Indels:         0
DB:                     14              Gaps:           0

US-09-980-881A-4 (1-246) x CB156851 (1-510)

Qy         85 ArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeu 104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGACCTG 60

Qy        105 AsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSer 124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTCG 120

Qy        125 GluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPhe 144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGTTTC 180

Qy        145 LeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHis 164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCAT 240

Qy        165 IleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeu 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCTA 300

Qy        185 ValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHis 204
              ||||||||||||||||||||||||||||||||  ||||||||||||||||||||||||||||||
Db        301 GTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACAT 360

Qy        205 GlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAsp 224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTATGAT 420

Qy        225 LeuGlyIleLysTyrSerPheThr 232
              |||||||||||||||||||||||
Db        421 TTGGGCATCAAATATTCGTTTACA 444

RESULT 8
AL536069
LOCUS       AL536069                971 bp    mRNA    linear   EST 31-MAY-2003
DEFINITION  AL536069 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF022YE21 5-PRIME, mRNA sequence.
ACCESSION   AL536069
VERSION     AL536069.2  GI:31260939
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 971)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12799562.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6944.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DF022AC11QP1&cluster=6944.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF022AC11QP1.
FEATURES              Location/Qualifiers
     source           1. .971
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CS0DF022YE21"
                      /tissue_type="FETAL BRAIN"
                      /dev_stage="fetal"
                      /clone_lib="Homo sapiens FETAL BRAIN"
                      /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                      was primed with a NotI-oligo(dT) primer. Five prime end
                      enriched, double-strand cDNA was digested with Not I and
                      cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                      vector. Library was not normalized."
BASE COUNT      282 a    201 c    211 g    272 t      5 others
ORIGIN

Alignment Scores:
Pred. No.:              9.05e-80        Length:         971
Score:                  793.50          Matches:        148
Percent Similarity:     79.14%          Conservative:   0
Best Local Similarity:  79.14%          Mismatches:     2
Query Match:            59.30%          Indels:         37
DB:                     9               Gaps:           1

US-09-980-881A-4 (1-246) x AL536069 (1-971)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        407 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 466

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        467 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 526

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db        527 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 586

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        587 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 646

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        647 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 706

Qy         83 ---------------------------------------------------------- 83

Db        707 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAG 766

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        767 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATTGGAACAGAC 826

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        827 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 886

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        887 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 946

Qy        144 PheLeuArgArgAsnIleAsn 150
              ||||||||||||  ||||||
Db        947 TTCTTGAGAAGAAAWATMAAC 967

RESULT 9
CB164014
LOCUS       CB164014                434 bp    mRNA    linear   EST 30-JAN-2003
DEFINITION  K-EST0225031 L17N670205n1 Homo sapiens cDNA clone
            L17N670205n1-38-G03 5', mRNA sequence.
ACCESSION   CB164014
VERSION     CB164014.1  GI:28150140
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
```

```
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 434)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 38  row: G  column: 03
            High quality sequence stop: 434.
FEATURES             Location/Qualifiers
     source          1. .434
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L17N670205n1-38-G03"
                     /sex="F"
                     /lab_host="Top10F'"
                     /clone_lib="L17N670205n1"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
BASE COUNT      133 a     92 c    100 g    109 t
ORIGIN

Alignment Scores:
Pred. No.:              2.72e-73        Length:         434
Score:                  732.00          Matches:        137
Percent Similarity:     96.53%          Conservative:   2
Best Local Similarity:  95.14%          Mismatches:     4
Query Match:            54.71%          Indels:         1
DB:                     14              Gaps:           0

US-09-980-881A-4 (1-246) x CB164014 (1-434)

Qy         65 GlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHisAsn 84
              ||||||||||||||||||||||||||||||||||||||||||||||||||   |||||||||:::
Db          3 GGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTGCATAGGCCATAGT 62

Qy         85 ArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThr-AspLe 104
              ||||||||||||||||||||:::|||||||||||||||||||||||||||||| |||||
Db         63 CGAATGTGGAGAAAGAACCGGACTTTCTATGCGAACAATCATTGCATCGGAACAGGACCT 122

Qy        104 uAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSe 124
              ||||||||||||||||||||    ||||||||||||||||||||||||||||||||||||
Db        123 GAATAGGAACTTTGCTTCCAAACGCTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGCTC 182

Qy        124 rGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSerPh 144
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db        183 AGAAACCTACTGTGGACTTTATCCTGAGTCGTAACCAGAAGTGAAGGCAGTGGCTAGTTT 242

Qy        144 eLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHi 164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 CTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAGCA 302

Qy        164 sIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSerLe 184
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        303 TATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCTCT 362

Qy        184 uValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHi 204
              ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db        363 AGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACCAGGTATACACA 422

Qy        204 sGlyHisGly 207
              |||||||||
Db        423 TGGCCATGGC 432

RESULT 10
AW413090/c
LOCUS       AW413090                723 bp    mRNA    linear   EST 08-FEB-2000
DEFINITION  uq52c03.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:2922052 3' similar to TR:Q15114 Q15114 PREPRO-PLASMA
            CARBOXYPEPTIDASE B. ;, mRNA sequence.
ACCESSION   AW413090
VERSION     AW413090.1  GI:6939027
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 723)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1047144
            Seq primer: custom primer used
            High quality sequence stop: 494.
FEATURES             Location/Qualifiers
     source          1. .723
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2922052"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse liver mlia"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      191 a    153 c    159 g    220 t
ORIGIN

Alignment Scores:
Pred. No.:              4.07e-69        Length:         723
Score:                  698.50          Matches:        139
Percent Similarity:     81.22%          Conservative:   8
Best Local Similarity:  76.80%          Mismatches:     16
Query Match:            52.20%          Indels:         18
DB:                     9               Gaps:           1

US-09-980-881A-4 (1-246) x AW413090 (1-723)
```

```
Qy          84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
               ||||||     ||||||||||||||    :::    ||||||    |||:::|||||||||
Db         685 AATCGAAAGTGAAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAC 626

Qy         104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
               ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db         625 TTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 566

Qy         124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         565 TCTGAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 506

Qy         144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
               ||||||||||||||||||:::   ||||||||||||||||||||||||||||||||||||
Db         505 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAA 446

Qy         164 HisIleValProProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
               |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db         445 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCT 386

Qy         184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
               ||||||||||||||||||||||||||||||||||      :::||||||||||||||||||
Db         385 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACA 326

Qy         204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
               ||||||   |||||||||:::||||||||||||||||||||   ||||||||||||||||
Db         325 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 266

Qy         224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
               |||||||||||||||||||||||
Db         265 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 206

Qy         232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||   |||:::||||||   |||   ||||||||
Db         205 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 146

Qy         246 s 246
               |
Db         145 A 145
```

RESULT 11
AI255929/c

```
LOCUS       AI255929                772 bp    mRNA    linear   EST 12-NOV-1998
DEFINITION  ui83d11.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1889013 3' similar to TR:Q15114 Q15114 PREPRO-PLASMA
            CARBOXYPEPTIDASE B. ;, mRNA sequence.
ACCESSION   AI255929
VERSION     AI255929.1  GI:3863454
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 772)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:973337
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 477.
FEATURES             Location/Qualifiers
     source          1. .772
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1889013"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse liver mlia"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      204 a    158 c    167 g    238 t      5 others
ORIGIN
```

Alignment Scores:

Pred. No.:	1.38e-67	Length:	772
Score:	685.50	Matches:	141
Percent Similarity:	75.12%	Conservative:	10
Best Local Similarity:	70.15%	Mismatches:	24
Query Match:	51.23%	Indels:	26
DB:	9	Gaps:	3

US-09-980-881A-4 (1-246) x AI255929 (1-772)

```
Qy          67 HisAla---------ArgGluTrpIleSerProAlaPheCysLeuTrpPheIleGlyHis 83
               ||||||         ||||||||:::                          :::
Db         761 CATGCCCGTGATGAACGTGAATGGCTA--------------TGACTACACGTGGAAAAG 717

Qy          84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
               ||||||||||||||||||   ||||||   :::   ||||||   |||:::|||||||||
Db         716 AATCGAATGTGGAGAAAGANCCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAC 657

Qy         104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
               ||||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db         656 TTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 597

Qy         124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
               ||||||||||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db         596 TCTGAAACCTACTGTGGACTTTATCNTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 537

Qy         144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
               ||||||||||||||||||:::   |||   ||||||||||||||||||||||||||||||
Db         536 TTCTTGAGAAGAAATATCGACCACATTANAGCTTACATCAGTATGCACTCATACTCCCAA 477

Qy         164 HisIleValProProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
               |||:::|||||||||||||||   ||||||||||||||         ||||||||||||
Db         476 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGNCCNCGAAGAACTGTCT 417

Qy         184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
               ||||||||||||||||||||||||||||||||||||||      :::||||||||||||||||||
Db         416 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACA 357

Qy         204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
               ||||||   |||||||||:::||||||||||||||||||||   ||||||||||||||||
Db         356 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 297

Qy         224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
               |||||||||||||||||||||||
```

```
Db        296 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 237

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                              ||||||||||||   |||:::||||||   |||   ||||||||
Db        236 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 177

Qy        246 s 246
                |
Db        176 A 176
```

RESULT 12
BF384322

```
LOCUS       BF384322                692 bp    mRNA    linear   EST 27-NOV-2000
DEFINITION  602046670F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:4196129 5',
            mRNA sequence.
ACCESSION   BF384322
VERSION     BF384322.1  GI:11365627
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 692)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9532  row: 1  column: 18
            High quality sequence stop: 689.
FEATURES             Location/Qualifiers
     source          1. .692
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4196129"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Li9"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      211 a    156 c    139 g    186 t
ORIGIN
```

Alignment Scores:

Pred. No.:	1.24e-66	Length:	692
Score:	676.50	Matches:	134
Percent Similarity:	81.61%	Conservative:	8
Best Local Similarity:	77.01%	Mismatches:	14
Query Match:	50.56%	Indels:	18
DB:	10	Gaps:	1

US-09-980-881A-4 (1-246) x BF384322 (1-692)

```
Qy         91 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAspLeuAsnArgAsnPheAlaSer 110
              ||||||   :::   ||||||   |||:::||||||||||||||||||||||||||||||
Db          1 CGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGACCTGAACAGGAACTTCGCTTCC 60

Qy        111 LysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLeu 130
              ||||||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGACTT 120

Qy        131 TyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAsn 150
              ||||||||||||||||||||||||||||||||||   ||||||||||||||||||:::
```

```
Db        121 TATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATCGAC 180

Qy        151 GlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValProTyrSer 170
              ||||||||||||||||||||||||||||||||||||||||||   |||:::||||||||||||
Db        181 CACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTATTCC 240

Qy        171 TyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVal 190
              |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCAGTT 300

Qy        191 ArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluThr 210
              ||||||||||||      :::||||||||||||||||||||||||   |||||||||:::
Db        301 CGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAAAGT 360

Qy        211 LeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSer 230
              ||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        361 TTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATATTCG 420

Qy        231 Phe----------------------------------------------------ThrSe 233
              |||                                                    |||||
Db        421 TTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATACATC 480

Qy        233 rAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
              |||||||   |||:::||||||   |||   |||||||||
Db        481 AAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 520
```

RESULT 13
AI255397/c

```
LOCUS       AI255397                747 bp    mRNA    linear   EST 12-NOV-1998
DEFINITION  ui94a04.x1 Sugano mouse liver mlia Mus musculus cDNA clone
            IMAGE:1890030 3' similar to TR:Q15114 Q15114 PREPRO-PLASMA
            CARBOXYPEPTIDASE B. ;, mRNA sequence.
ACCESSION   AI255397
VERSION     AI255397.1  GI:3862922
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 747)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:974354
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 471.
FEATURES             Location/Qualifiers
     source          1. .747
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1890030"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse liver mlia"
                     /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
```

was primed with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACA."

BASE COUNT 199 a 151 c 162 g 233 t 2 others
ORIGIN

Alignment Scores:

Pred. No.:	1.22e-64	Length:	747
Score:	659.50	Matches:	132
Percent Similarity:	77.90%	Conservative:	9
Best Local Similarity:	72.93%	Mismatches:	22
Query Match:	49.29%	Indels:	18
DB:	9	Gaps:	1

US-09-980-881A-4 (1-246) x AI255397 (1-747)

```
Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              |||||||||||||||||           :::      |||     :::|||||||||
Db        711 AATCGAATGTGGAGGAAGACCGCTCTTGCTCACAAGACCAACCGTTGNGTGGGCACAGAC 652

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              |||   ||||||||||||   ||||||||||||||:::|||||||||||||||||||||
Db        651 CTGACCAGGAACTTTGCTTTCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 592

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              |||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        591 TCTAAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 532

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||:::   ||||||||||||||||||||||||||||||||||||
Db        531 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAA 472

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 |||:::|||||||||||||||   ||||||||||||||||||   ||||||||||||
Db        471 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCNCGAAGAACTGTCT 412

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              |||||||||||||||||||||||||||||||||||||      :::|||||||||||||||
Db        411 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACA 352

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||   |||||||||:::|||||||||||||||||||||   ||||||||||||||
Db        351 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 292

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db        291 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 232

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||   |||:::||||||   |||   |||||||
Db        231 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 172

Qy        246 s 246
              |
Db        171 A 171
```

RESULT 14
BI246210
LOCUS BI246210 713 bp mRNA linear EST 17-JUL-2001
DEFINITION 602958724F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5124700 5', mRNA sequence.
ACCESSION BI246210
VERSION BI246210.1 GI:14789700
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 713)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11304 row: o column: 05
High quality sequence stop: 686.
FEATURES Location/Qualifiers
source 1. .713
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5124700"
/lab_host="DH10B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_Li9"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 212 a 168 c 145 g 188 t
ORIGIN

Alignment Scores:

Pred. No.:	2.5e-64	Length:	713
Score:	656.50	Matches:	132
Percent Similarity:	81.14%	Conservative:	10
Best Local Similarity:	75.43%	Mismatches:	14
Query Match:	49.07%	Indels:	19
DB:	12	Gaps:	1

US-09-980-881A-4 (1-246) x BI246210 (1-713)

```
Qy         91 ArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp-LeuAsnArgAsnPheAlaSe 110
              ||||||   :::   ||||||   |||:::|||||::: :::|||||||||||||||||
Db          2 CGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAGCGTGAACAGGAACTTCGCTTC 61

Qy        110 rLysHisTrpCysGluGluGlyAlaSerSerSerSerCysSerGluThrTyrCysGlyLe 130
              ||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db         62 CAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGCTCTGAAACCTACTGTGGACT 121

Qy        130 uTyrProGluSerGluProGluValLysAlaValAlaSerPheLeuArgArgAsnIleAs 150
              ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||::
Db        122 TTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGACTTCTTGAGAAGAAATATCGA 181

Qy        150 nGlnIleLysAlaTyrIleSerMetHisSerTyrSerGlnHisIleValPheProTyrSe 170
              :   ||||||||||||||||||||||||||||||||||   |||:::||||||||||||
Db        182 CCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAACAAATACTGTTTCCCTATTC 241

Qy        170 rTyrThrArgSerLysSerLysAspHisGluGluLeuSerLeuValAlaSerGluAlaVa 190
              ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        242 CTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCTCTAGTGGCCAGCGAAGCAGT 301

Qy        190 lArgAlaIleGluLysThrSerLysAsnThrArgTyrThrHisGlyHisGlySerGluTh 210
              ||||||||||||      :::|||||||||||||||||||||||   |||||||||::
Db        302 TCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACACACGGCAGTGGCTCAGAAAG 361

Qy        210 rLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyrAspLeuGlyIleLysTyrSe 230
              :||||||||||||||||||   ||||||||||||||||||||||||||||||||||||||
Db        362 TTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTATGATTTGGGCATCAAATATTC 421

Qy        230 rPhe----------------------------------------------------ThrS 233
```

```
              ||||                                                    ||||
Db        422 GTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTCTTGCTGCCTGAGAGATACAT 481

Qy        233 erAsnProProValGluLysLeuLeuProLeuSerLeuLys 246
              ||||||||   |||:::||||||   |||   |||||||||
Db        482 CAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAAA 522

RESULT 15
CB161965
LOCUS       CB161965               505 bp    mRNA    linear   EST 30-JAN-2003
DEFINITION  K-EST0222418 L17N670205n1 Homo sapiens cDNA clone
            L17N670205n1-15-E06 5', mRNA sequence.
ACCESSION   CB161965
VERSION     CB161965.1  GI:28148091
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
  TITLE     21C Frontier Korean EST Project 2001
  JOURNAL   Unpublished
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 15  row: E  column: 06
            High quality sequence stop: 505.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="L17N670205n1-15-E06"
                     /sex="F"
                     /lab_host="Top10F'"
                     /clone_lib="L17N670205n1"
                     /note="Organ: Liver; Vector: pT7T3-Pac; Site_1: EcoRI;
                     Site_2: NotI; The library was contributed by the Soares
                     laboratory and it was constructed as described by Bonaldo,
                     M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
                     6(9): 791-806. RNA was prepared from harvested cell
                     culture."
BASE COUNT      160 a     99 c    108 g    138 t
ORIGIN

Alignment Scores:
Pred. No.:              4.18e-64        Length:         505
Score:                  652.50          Matches:        125
Percent Similarity:     75.45%          Conservative:   1
Best Local Similarity:  74.85%          Mismatches:     4
Query Match:            48.77%          Indels:         37
DB:                     14              Gaps:           1

US-09-980-881A-4 (1-246) x CB161965 (1-505)

Qy         78 LeuTrpPheIleGlyHis---------------------------------------- 83
              ||||||||||||||||||
Db          3 TTGTGGTTCATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTC 62

Qy         83 ------------------------------------------------------------ 83

Db         63 CTGAGGCTTGTGGATTTCTATGTTATGCCAGTGGTTAATGTGGATGGTTATGACTACTCA 122

Qy         84 ---------AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
                      |||  ||||||||||||||||||||||||||||||||||||||||||||||
Db        123 TGGAAAAAGAATCTAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 182

Qy        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        183 GGAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 242

Qy        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        243 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 302

Qy        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
              ||||||||||||||||   ||||||||||||||||||||||||||||||   |||||||||
Db        303 GTGGCTAGTTTCTTGACAAGAAATATCAACCAGATTAAAGCATACATCATCATGCATTCA 362

Qy        161 TyrSerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        363 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 422

Qy        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
              ||||||||||||:::||||||||||||||||||||||||||||||   ||||||||||||
Db        423 GAACTGTCTCTAATAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACC 482

Qy        201 ArgTyrThrHisGlyHisGly 207
              |||||||||||||||||||||
Db        483 AGGTATACACATGGCCATGGC 503


Search completed: January  2, 2004, 02:37:40
Job time : 2189.08 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:00:56 ; Search time 42.8831 Seconds
(without alignments)
910.539 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1338	100.0	246	21	AAB11459	Human brain carbox
2	1338	100.0	338	21	AAB11458	Human brain carbox
3	1325	99.0	360	21	AAB11457	Human brain carbox
4	1262	94.3	386	24	ABG72095	Human protease pro
5	1257	93.9	386	24	ABG72093	Human protease pro
6	1239.5	92.6	423	17	AAR90293	Human plasma carbo
7	1239.5	92.6	423	18	AAW14733	Human plasma carbo
8	1239.5	92.6	423	20	AAW92270	Human plasma carbo
9	1234.5	92.3	423	14	AAR36273	Human plasma carbo
10	816.5	61.0	211	21	AAB58129	Lung cancer associ
11	612.5	45.8	324	22	AAG66562	Bothops jararaca c
12	610	45.6	307	17	AAW00602	Rat carboxypeptida
13	582	43.5	306	16	AAR75132	Porcine carboxypep
14	582	43.5	306	22	AAU04477	Porcine carboxypep
15	582	43.5	404	16	AAR75131	Porcine Tyr-His-Me
16	567	42.4	329	17	AAW06174	Mature HCPB with P
17	567	42.4	349	17	AAW06173	mature HCPB-(His)6
18	567	42.4	415	17	AAW06172	Human pancreatic c
19	567	42.4	415	19	AAW74476	Human pancreatic c
20	567	42.4	417	24	ABB99688	Amino acid sequenc
21	567	42.4	424	17	AAW06175	ProHCPB with PelB
22	560	41.9	349	17	AAW06181	Modified HCPB (D25
23	560	41.9	349	18	AAW13749	Carboxypeptidase B
24	559	41.8	349	17	AAW06182	Modified HCPB (D25
25	559	41.8	349	18	AAW13750	Carboxypeptidase B
26	556	41.6	349	18	AAW13763	Carboxypeptidase B
27	554	41.4	349	18	AAW13752	Carboxypeptidase B
28	554	41.4	349	18	AAW13758	Carboxypeptidase B
29	553	41.3	349	18	AAW13751	Carboxypeptidase B
30	552	41.3	349	18	AAW13753	Carboxypeptidase B
31	552	41.3	417	23	AAU84325	Protein CPA3 diffe
32	552	41.3	667	24	ABP97414	Anti-human seminal
33	551	41.2	311	22	AAG66563	Human carboxypepti
34	551	41.2	349	18	AAW13760	Carboxypeptidase B
35	551	41.2	349	18	AAW13761	Carboxypeptidase B
36	551	41.2	349	18	AAW13762	Carboxypeptidase B
37	550.5	41.1	354	22	AAG66560	Human secreted met
38	550.5	41.1	354	22	AAG66566	Human secreted met
39	550.5	41.1	362	22	AAB74682	Human protease and
40	550.5	41.1	374	22	AAG66547	Human secreted met
41	550.5	41.1	374	22	AAG66565	Human secreted met
42	550.5	41.1	374	23	AAU82703	Amino acid sequenc
43	550.5	41.1	374	24	ABJ37888	NOVX protein seque
44	550.5	41.1	374	24	ABG74658	Human zinc-binding
45	550.5	41.1	444	23	AAE24240	Human 23566 (carbo

ALIGNMENTS

RESULT 1
AAB11459
ID AAB11459 standard; Protein; 246 AA.
XX
AC AAB11459;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 4.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
XX

PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Disclosure; Page 74-75; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 246 AA;

Query Match 100.0%; Score 1338; DB 21; Length 246;
Best Local Similarity 100.0%; Pred. No. 5.8e-134;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240

Qy        241 LPLSLK 246
              ||||||
Db        241 LPLSLK 246
```

RESULT 2
AAB11458
ID AAB11458 standard; Protein; 338 AA.
XX
AC AAB11458;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein SEQ ID NO 3.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX
PI Matsumoto A;
XX
DR WPI; 2000-687534/67.
XX
PT Human brain carboxypeptidase B isolated from the hippocampus useful for
PT screening agents for the treatment of Alzheimer's and other brain
PT disorders -
XX
PS Disclosure; Page 71-73; 84pp; Japanese.
XX
CC This invention describes a novel protein with peptidase activity
CC against brain beta-amyloid precursor protein which has been isolated from
CC human hippocampus and which has cerebroprotective, antialzheimers,
CC nootropic, neuroprotective and hemostatic activity and which can be used
CC as a vaccine or for gene therapy. The protein, and compounds identified
CC by screening as promoters or inhibitors of its activity, are used to
CC regulate beta-amyloid accumulation in the brain and treat or prevent
CC diseases in which this occurs, such as Alzheimer's, senile dementia,
CC inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ Sequence 338 AA;

Query Match 100.0%; Score 1338; DB 21; Length 338;
Best Local Similarity 100.0%; Pred. No. 9.5e-134;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         93 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 152

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        153 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 212

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        213 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 272

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        273 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 332

Qy        241 LPLSLK 246
              ||||||
Db        333 LPLSLK 338
```

RESULT 3
AAB11457
ID AAB11457 standard; Protein; 360 AA.
XX
AC AAB11457;
XX
DT 01-MAR-2001 (first entry)
XX
DE Human brain carboxypeptidase B protein.
XX
KW Carboxypeptidase B; human; brain; hippocampus; APP cleavage; peptidase;
KW treatment; Alzheimer's disease; beta-amyloid precursor protein; vaccine;
KW cerebroprotective; antialzheimers; nootropic; neuroprotective;
KW hemostatic; gene therapy; senile dementia; inherited cerebral hemorrhage;
KW Down's syndrome; head trauma.
XX
OS Homo sapiens.
XX
PN WO200066717-A1.
XX
PD 09-NOV-2000.
XX
PF 01-MAY-2000; 2000WO-JP02878.
XX
PR 30-APR-1999; 99JP-0125169.
XX
PA (MATS/) MATSUMOTO A.
XX

```
PI   Matsumoto A;
XX
DR   WPI; 2000-687534/67.
DR   N-PSDB; AAC81962.
XX
PT   Human brain carboxypeptidase B isolated from the hippocampus useful for
PT   screening agents for the treatment of Alzheimer's and other brain
PT   disorders -
XX
PS   Claim 1; Page 68-71; 84pp; Japanese.
XX
CC   This invention describes a novel protein with peptidase activity
CC   against brain beta-amyloid precursor protein which has been isolated from
CC   human hippocampus and which has cerebroprotective, antialzheimers,
CC   nootropic, neuroprotective and hemostatic activity and which can be used
CC   as a vaccine or for gene therapy. The protein, and compounds identified
CC   by screening as promoters or inhibitors of its activity, are used to
CC   regulate beta-amyloid accumulation in the brain and treat or prevent
CC   diseases in which this occurs, such as Alzheimer's, senile dementia,
CC   inherited cerebral hemorrhage, Down's syndrome, and head trauma.
XX
SQ   Sequence   360 AA;

  Query Match            99.0%;  Score 1325;  DB 21;  Length 360;
  Best Local Similarity  98.8%;  Pred. No. 2.5e-132;
  Matches  243;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              |||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNSNFVSKHWCEEGASS 234

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 294

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240
              |||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db        295 ELSLVASEAVRAIDKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 354

Qy        241 LPLSLK 246
              ||||||
Db        355 LPLSLK 360


RESULT 4
ABG72095
ID   ABG72095 standard; Protein; 386 AA.
XX
AC   ABG72095;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #2.
XX
KW   Human; enzyme; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 169
FT                   /note= "This residue is the minor variant of a single
FT                         nucleotide polymorphism. The major variant is an
FT                         Ala residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Page -; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #2, the gene for which is located on chromosome 13.
CC   Note: This sequence is not shown in the specification but has been
CC   created by the indexer from information given in figure 3.
XX
SQ   Sequence   386 AA;

  Query Match            94.3%;  Score 1262;  DB 24;  Length 386;
  Best Local Similarity  99.6%;  Pred. No. 1.5e-125;
  Matches  231;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 234

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 294

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        295 ELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
```

```
RESULT 5
ABG72093
ID   ABG72093 standard; Protein; 386 AA.
XX
AC   ABG72093;
XX
DT   14-FEB-2003  (first entry)
XX
DE   Human protease protein, variation #1.
XX
KW   Human; enzyme; protease; proteolytic degradation; proteolysis;
KW   proliferation; differentiation; signalling; therapeutic; gene therapy;
KW   protein therapy; diagnostic; immune response; vaccine; inflammation;
KW   cancer; arteriosclerosis; degenerative disorder; chromosome 13;
KW   single nucleotide polymorphism; SNP.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 169
FT                   /note= "This residue is the major variant of a single
FT                         nucleotide polymorphism. The minor variant is an
FT                         Thr residue"
XX
PN   US2002137179-A1.
XX
PD   26-SEP-2002.
XX
PF   21-MAR-2001; 2001US-0813133.
XX
PR   21-MAR-2001; 2001US-0813133.
XX
PA   (GANW/) GAN W.
PA   (KETC/) KETCHUM K A.
PA   (DFRA/) DI FRANCESCO V.
PA   (BEAS/) BEASLEY E M.
XX
PI   Gan W,  Ketchum KA,  Di Francesco V,  Beasley EM;
XX
DR   WPI; 2003-102387/09.
DR   N-PSDB; ABX13670, ABX13671.
XX
PT   New isolated human protease proteins, useful for developing therapeutic
PT   or diagnostic compositions, particularly for developing human
PT   therapeutic agents that modulate protease activity in cells or tissues
PT   -
XX
PS   Claim 1; Fig 2; 80pp; English.
XX
CC   The invention discloses an isolated human protease peptide, its allelic
CC   variant or orthologue. The proper functioning of the cell requires
CC   careful control of the levels of important structural proteins, enzymes
CC   and regulatory proteins. One of the ways the cell reduces the steady
CC   state level of a particular protein is by proteolytic degradation.
CC   Proteolysis can also be used to convert a pre or pro-protein in to an
CC   active form. Proteases also regulate many different cell proliferation,
CC   differentiation and signalling processes. The peptides and nucleic acid
CC   molecules are useful in the development of human therapeutics (gene and
CC   protein therapy) and diagnostic compositions. The peptides are also
CC   useful for raising antibodies or eliciting an immune response (vaccine),
CC   as a reagent (including the labeled reagent) in assays designed to
CC   quantitatively determine levels of the protein (or its binding partner or
CC   ligand) in biological fluids, or as markers for tissues in which the
CC   corresponding protein is preferentially expressed and in methods for
CC   identifying a modulator of the peptide or an agent that binds to the
CC   peptide. The agents identified are useful for treating protease-related
CC   conditions that are specific for the subfamily of proteases that the
CC   peptide belongs to, particularly in cells and tissues that express the
CC   protease, such as inflammation, cancer, arteriosclerosis and degenerative
CC   disorders. The modulator of the peptide is also useful for treating a
CC   disorder characterised by an absence of, inappropriate or unwanted
CC   expression of the protein. The sequence presented is the human protease
CC   protein, variant #1, the gene for which is located on chromosome 13.
XX
SQ   Sequence   386 AA;

  Query Match             93.9%;  Score 1257;  DB 24;  Length 386;
  Best Local Similarity   99.1%;  Pred. No. 5e-125;
  Matches  230;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 234

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 294

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        295 ELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346


RESULT 6
AAR90293
ID   AAR90293 standard; Protein; 423 AA.
XX
AC   AAR90293;
XX
DT   25-MAR-2003  (updated)
DT   12-APR-1996  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Plasma carboxypeptidase B; hPCPB; antibody; detection;
KW   purification; plasminogen; affinity column.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Active-site     181
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    183
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    256..257
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     310
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    312
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
FT   Active-site     369
FT                   /note= "Residue suspecting of being involved in
FT                         catalytic activity of enzyme."
FT   Binding-site    370..371
FT                   /note= "Residues suspected of being part of the
FT                         binding site of enzyme."
FT   Binding-site    394
FT                   /note= "Residue suspected of being part of the
FT                         binding site of enzyme."
XX
PN   US5474901-A.
XX
PD   12-DEC-1995.
XX
PF   19-JUL-1994;   94US-0277540.
XX
```

```
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1996-039508/04.
DR   N-PSDB; AAT11671.
XX
PT   Antibody to human plasma carboxy:peptidase B - useful for detecting
PT   and purifying hPCPB for use in treating clotting disorders e.g.
PT   haemophilia A
XX
PS   Disclosure; Figure 4; 40pp; English.
XX
CC   An antibody which specifically binds human plasma carboxypeptidase B
CC   (hPCPB) and does not cross react with other carboxypeptidases is
CC   useful for the detection of hPCPB in vitro.  The antibody is also
CC   used for purifying hPCPB from a sample.  Purification comprises
CC   passing a sample thought to contain hPCPB over either a column to
CC   which antibody has been bound, or a plasminogen affinity column,
CC   eluting the column and then recovering the fraction containing the
CC   hPCPB.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

  Query Match             92.6%;  Score 1239.5;  DB 17;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 4.2e-123;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH----------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 7
AAW14733
ID   AAW14733 standard; Protein; 423 AA.
XX
AC   AAW14733;
XX
DT   25-MAR-2003  (updated)
DT   08-MAY-1997  (first entry)
XX
DE   Human plasma carboxypeptidase B.
XX
KW   Human; plasma carboxypeptidase B; PCPB; haemostatic regulation;
KW   plasma; plasminogen.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..22
FT                   /note= "Signal peptide"
FT   Protein         23..423
FT                   /note= "Mature PCPB"
FT   Cleavage-site   114
FT                   /note= "Potential clip site for activation of PCPB
FT                        as a carboxypeptidase"
FT   Active-site     181
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    183
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     184
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    256
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    257
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     310
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    312
FT                   /note= "Expected to be involved in substrate binding"
FT   Active-site     369
FT                   /note= "Expected to be involved in catalytic activity"
FT   Binding-site    370
FT                   /note= "Expected to be involved in substrate
FT                        binding, thought to determine specificity
FT                        of PCPB as a carboxypeptidase B"
FT   Binding-site    371
FT                   /note= "Expected to be involved in substrate binding"
FT   Binding-site    394
FT                   /note= "Expected to be involved in substrate binding"
XX
PN   US5593674-A.
XX
PD   14-JAN-1997.
XX
PF   27-APR-1995;   95US-0430787.
XX
PR   01-FEB-1991;   91US-0649591.
PR   14-OCT-1992;   92US-0959944.
PR   15-DEC-1993;   93US-0167727.
PR   19-JUL-1994;   94US-0277540.
PR   27-APR-1995;   95US-0430787.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1997-099413/09.
DR   N-PSDB; AAT62846.
XX
PT   Using human plasma carboxy:peptidase B in blood coagulation - is
PT   functionally related to carboxy:peptidase A and pancreas
PT   carboxy:peptidase B
XX
PS   Example 2; Column 37-42; 39pp; English.
XX
CC   This sequence represents human plasma carboxypeptidase B (PCPB) which
CC   has a molecular weight under non-reducing SDS-PAGE of approx. 60 kD.
CC   PCPB may be used therapeutically in haemostatic regulation.  PCPB is
CC   purified from human plasma or by transformed cell culture by
CC   extraction using plasminogen bound to a solid phase.
CC   (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ   Sequence   423 AA;

  Query Match             92.6%;  Score 1239.5;  DB 18;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 4.2e-123;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174
```

```
Qy          61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
               |||||||||||||||||||||||
Db         175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy          84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy         144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy         204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
               |||||||||||||||||||||||||||||
Db         355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383
```

RESULT 8
AAW92270
ID AAW92270 standard; Protein; 423 AA.
XX
AC AAW92270;
XX
DT 28-APR-1999 (first entry)
XX
DE Human plasma carboxypeptidase B (PCPB) thr147.
XX
KW Plasma carboxypeptidase B; PCPB; human; hPCPBthr147;
KW polymorphism detection; thrombotic disease.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT		/note= "signal peptide"
FT	Protein	23..423
FT		/note= "mature PCPBthr147"

XX
PN WO9855645-A1.
XX
PD 10-DEC-1998.
XX
PF 02-JUN-1998; 98WO-EP03244.
XX
PR 03-JUN-1997; 97US-0869057.
XX
PA (SCHD) SCHERING AG.
XX
PI Morser MJ, Nagashima M;
XX
DR WPI; 1999-045800/04.
DR N-PSDB; AAV74302.
XX
PT Detecting new polymorphism of human plasma carboxypeptidase B -
PT comprises Alanine or Threonine at position 147 of protein by DNA or
PT protein analysis, useful to detect risk of thrombotic disease in
PT humans
XX
PS Disclosure; Page 25-26; 35pp; English.
XX
CC This sequence is the human plasma carboxypeptidase B (PCPB) mutant
CC hPCPBthr147. The invention relates to a method for determining the
CC presence of DNA or protein polymorphisms of PCPB in human subjects, which
CC comprises obtaining a prepared tissue or blood sample and determining the
CC presence of DNA coding for naturally occurring polymorphs of the protein
CC containing Alanine or Threonine at position 147 (PCPB1 and PCPB2
CC respectively). Determination of the relative distribution of the PCPB
CC polymorphs in a patient's blood by genetic or protein analysis by the
CC methods is useful to determine the risk of thrombotic disease in humans.
CC Such assessments may be made by accumulating information concerning the
CC relative distribution of the different polymorphs within the general
CC population compared with populations known to be at risk and establishing
CC a PCPB polymorph profile for at-risk patients.
XX
SQ Sequence 423 AA;

Query Match 92.6%; Score 1239.5; DB 20; Length 423;
Best Local Similarity 86.2%; Pred. No. 4.2e-123;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy           1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy          61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
               |||||||||||||||||||||||
Db         175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy          84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy         144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy         204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
               |||||||||||||||||||||||||||||
Db         355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383
```

RESULT 9
AAR36273
ID AAR36273 standard; Protein; 423 AA.
XX
AC AAR36273;
XX
DT 24-AUG-1993 (first entry)
XX
DE Human plasma carboxypeptidase B.
XX
KW PCPB; tissue plasminogen activator inhibitor; t-PA inhibitor.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	Peptide	1..22
FT		/label= signal_peptide
FT	Cleavage-site	114
FT		/note= "putative site of cleavage by trypsin to
FT		activate PCPB as a carboxypeptidase"
FT	Active-site	181
FT		/note= "expected to be involved in catalytic
FT		activity"
FT	Binding-site	183
FT		/note= "expected to be involved in substrate
FT		binding"
FT	Active-site	184
FT		/note= "expected to be involved in catalytic
FT		activity"
FT	Binding-site	256..257
FT		/note= "expected to be involved in substrate
FT		binding"
FT	Active-site	310
FT		/note= "expected to be involved in catalytic
FT		activity"
FT	Binding-site	312
FT		/note= "expected to be involved in substrate
FT		binding"
FT	Active-site	369
FT		/note= "expected to be involved in catalytic
FT		activity"
FT	Binding-site	370..371
FT		/note= "expected to be involved in substrate
FT		binding; Asp 370 is the residue which
FT		determines specificity of the PCPB as

```
FT                              a carboxypeptidase B"
FT   Binding-site    394
FT                   /note= "expected to be involved in substrate
FT                          binding"
XX
PN   US5206161-A.
XX
PD   27-APR-1993.
XX
PF   01-FEB-1991;   91US-0649591.
XX
PR   01-FEB-1991;   91US-0649591.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Drayna DT,  Eaton DL;
XX
DR   WPI; 1993-151724/18.
DR   N-PSDB; AAQ41001.
XX
PT   New human plasma carboxypeptidase B - used as haemostatic
PT   regulator for clotting blood, partic. for treating blood clotting
PT   disorders, e.g. haemophilia
XX
PS   Claim 2; Fig 4; 40pp; English.
XX
CC   This amino acid sequence was deduced from a clone isolated from a
CC   human liver cDNA library. There is about 40% sequence identity
CC   between the deduced preproPCPB amino acid sequence and pro-human
CC   mast cell carboxypeptidase A and between preproPCPB and prepro-rat
CC   carboxypeptidase B. Human PCPB has the same substrate binding sites
CC   as, and shares 6 cysteine residues which form 3 intramolecular
CC   disulphide bonds with, bovine and rat carboxypeptidase B. The
CC   presence of the same amino acid (Asp at position 348 of the mature
CC   protein) at the region in carboxypeptidases that determines
CC   substrate specificity, suggests that PCPB represents a plasma-
CC   derived carboxypeptidase. PCPB inhibits the enzymatic conversion by
CC   tPA of plasminogen to plasmin in the presence of fibrinogen.
XX
SQ   Sequence   423 AA;

  Query Match             92.3%;  Score 1234.5;  DB 14;  Length 423;
  Best Local Similarity   85.9%;  Pred. No. 1.4e-122;
  Matches  231;  Conservative    0;  Mismatches    1;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH----------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGACSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 10
AAB58129
ID   AAB58129 standard; Protein; 211 AA.
XX
AC   AAB58129;
XX
DT   14-MAR-2001  (first entry)
```

```
XX
DE   Lung cancer associated polypeptide sequence SEQ ID 467.
XX
KW   Human; lung cancer associated protein; neuroprotective; cytostatic;
KW   cardioactive; immunomodulatory; muscular active; vulnerary;
KW   gastrointestinal; nephrotropic; antiinfective; gynecological;
KW   antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW   proliferative disorder; wound healing; infectious disease.
XX
OS   Homo sapiens.
XX
PN   WO200055180-A2.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05918.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (ROSE/) ROSEN C A.
XX
PI   Ruben SM;
XX
DR   WPI; 2000-587514/55.
DR   N-PSDB; AAF18005.
XX
PT   Lung cancer associated gene sequences, referred to as lung cancer
PT   antigens, useful for treatment, prevention, and diagnosis of disorders
PT   such as lung cancer -
XX
PS   Claim 11; Page 944-945; 1425pp; English.
XX
CC   Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
CC   associated proteins represented in AAB58106 - AAB58548. Lung cancer
CC   associated proteins and polynucleotide sequences, their agonists, and
CC   antagonists may have neuroprotective; cytostatic; cardioactive;
CC   immunomodulatory; muscular active general; vulnerary; gastrointestinal
CC   general; nephrotropic; antiinfective; gynecological; or antibacterial
CC   activity. The invention also includes antibodies specific for the
CC   protein or polynucleotide sequences. The lung cancer associated
CC   polynucleotide sequences may be used for detection of lung cancer,
CC   chromosome identification, as chromosome markers, and for numerous other
CC   diagnostic or research purposes. The proteins may be used to treat
CC   disorders such as neural, immune, muscular, reproductive,
CC   gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC   disorders. The proteins may also be used in the treatment of wounds and
CC   infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC   peptide AAB58549 are used in the course of the invention for the
CC   identification and characterisation of the polynucleotide and protein
CC   sequences.
XX
SQ   Sequence   211 AA;

  Query Match             61.0%;  Score 816.5;  DB 21;  Length 211;
  Best Local Similarity   90.6%;  Pred. No. 1.8e-78;
  Matches  154;  Conservative    1;  Mismatches    6;  Indels    9;  Gaps    2;

Qy         71 WISPAFCLWFI--------GHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 122
              |||   | | :          ||||||||||||||||||||||||||||||||||||||
Db          3 WIS-MLCRWLMWMVMNYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 61

Qy        123 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 121

Qy        183 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 171


RESULT 11
AAG66562
```

```
ID   AAG66562 standard; Protein; 324 AA.
XX
AC   AAG66562;
XX
DT   22-OCT-2001  (first entry)
XX
DE   Bothops jararaca carboxypeptidase homologue protein.
XX
KW   secreted metallocarboxypeptidase; immunomodulatory;
KW   antiparkinsonian; nootropic; anticonvulsant; neuroprotective;
KW   osteopathic; antiarthritic; cerebroprotective; vasotropic; vulnerary;
KW   antibacterial; antiviral; antifungal; antiinflammatory; antidiabetic;
KW   antirheumatic; cytostatic; antiallergic; haemostatic; infection;
KW   cancer.
XX
OS   Bothops jararaca.
XX
PN   WO200157265-A1.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03783.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
PR   29-SEP-2000; 2000US-0676135.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Boyle BJ,  Mize NK,  Arterburn MC,  Yeung G,  Tang YT,  Liu C;
PI   Drmanac RT,  Wang M,  Chen L,  Yang Y;
XX
DR   WPI; 2001-465705/50.
XX
PT   Novel metallocarboxy peptidase-like polypeptide is useful in treating
PT   neurodegenerative, immune, etc. disorders  -
XX
PS   Example 3; Fig 1; 146pp; English.
XX
CC   The invention relates to an isolated polypeptide comprising an amino
CC   acid sequence of 374, 41, 14, 15, 41, 22, 15, 27, 14, 9, 18, 20, 15,
CC   354, 165, 374, 354, or 41 amino acids, fully defined in the
CC   specification, or its translated protein coding portion, mature protein
CC   coding portion, extracellular portion or active domain. The polypeptide
CC   is a human secreted metallocarboxypeptidase-like polypeptide. It
CC   is useful for treating mammalian diseases including:
CC   neurodegenerative diseases such as Parkinson's or Alzheimer's diseases,
CC   osteoporosis, osteoarthritis, disorders linked to reduced tissue
CC   growth, Huntington's disease, amylotrophic lateral sclerosis, stroke,
CC   reperfusion, wound healing, infectious diseases, immune and autoimmune
CC   disorders such as severe combined immunodeficiency, rheumatoid
CC   arthritis, Guillain-Barre syndrome, insulin dependent diabetes
CC   mellitus, graft-versus-host disease and allergy, cancer, and
CC   haemophilia. The polypeptide is also useful in medical imaging, as an
CC   immunogen to generate antibodies, as a nutritional source or
CC   supplement, and as a contraceptive. The present sequence is
CC   a carboxypeptidase homologue protein of Bothops jararaca.
XX
SQ   Sequence   324 AA;

  Query Match             45.8%;  Score 612.5;  DB 22;  Length 324;
  Best Local Similarity   45.0%;  Pred. No. 2e-56;
  Matches  121;  Conservative   44;  Mismatches   63;  Indels   41;  Gaps    5;

Qy          2 SASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              :| | ||:|:|  :| :|   |  :| ::::: ||::||  |: :||| ||    | ||
Db         30 TAGYNYEKYNSWEKIDAWTADIANENPSLVSRLQIGTTFEGRPMPLLKV-GKPGVNKKAI 88

Qy         61 WIDCGIHAREWISPAFCLWFI-----------------------------------GH 83
              :|||| ||||||||||| ||:
Db         89 FIDCGFHAREWISPAFCQWFVREAVRTYGKETIMTQLLNKLDFYILPVLNIDGYVYSWKQ 148

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
```

```
              :||||| ||  | : ||||| |||| :  ||  ||| : |||||||  |||| | ||:|
Db        149 SRMWRKTRSVNAGSTCIGTDPNRNFDAA-WCSVGASRNPCSETYCGSKPESEKETKALAD 207

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:||| : |:||:::||||| :::|||||   : ::::|: :| ||:| : |    | ||
Db        208 FIRRNRSIIQAYLTIHSYSQMLLYPYSYTYDLTSNNKKLNSIAKEAIREL-KVLFGTEYT 266

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || ||||:||
Db        267 YGPGAATIYPAAGGSDDWAYDQGIKYAFT 295


RESULT 12
AAW00602
ID   AAW00602 standard; Protein; 307 AA.
XX
AC   AAW00602;
XX
DT   09-NOV-1996  (first entry)
XX
DE   Rat carboxypeptidase B mature enzyme.
XX
KW   Carboxypeptidase-B; pro-enzyme; protease; insulin.
XX
OS   Rattus sp.
XX
PN   WO9623064-A1.
XX
PD   01-AUG-1996.
XX
PF   25-JAN-1996;   96WO-US00995.
XX
PR   25-JAN-1995;   95US-0378233.
XX
PA   (BIOT-) BIOTECHNOLOGY GEN CORP.
XX
PI   Fulga N,  Gorecki M,  Hartman J,  Mendelovitch S;
XX
DR   WPI; 1996-362688/36.
DR   N-PSDB; AAT35760.
XX
PT   Purified active recombinant carboxypeptidase B prodn. - by
PT   expressing DNA encoding the pro-enzyme, folding and enzymatic
PT   cleavage to give active form, used e.g. for insulin prodn.
XX
PS   Example 1; Page 37-38; 49pp; English.
XX
CC   The amino acid sequence of the rat mature carboxypeptidase-B
CC   (CPB) (AAW00602) was deduced from a cDNA sequence (see also AAT35760)
CC   obtd. by PCR amplification of Sprague-Dawley rat pancreas cDNA.
CC   The sequence of the activation peptide (AAW00601) of pro-CPB has
CC   also been deduced.  Pro-CPB (pref. with a C290S mutation) can be
CC   expressed in host cells, e.g. Escherichia coli, and subsequently
CC   recovered, refolded and cleaved with trypsin to yield the active
CC   enzyme.  CPB produced this way is cheaper than porcine pancreatic
CC   enzyme, and is free of other proteases.
XX
SQ   Sequence   307 AA;

  Query Match             45.6%;  Score 610;  DB 17;  Length 307;
  Best Local Similarity   45.0%;  Pred. No. 3.3e-56;
  Matches  121;  Conservative   43;  Mismatches   65;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||   | :|::   | :||: :   :||::|:  ||::||   :||||: || :  | ||
Db          1 ASGHSYTKYNNWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAI 59

Qy         61 WIDCGIHAREWISPAFCLWFI-----------------------------GH------ 83
              :|||| |||||||||| ||:                             |:
Db         60 FIDCGFHAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTK 119

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :||||| ||  | : |:| | |||| :  ||| ||| | |||||||  |||| | ||:|
```

```
Db        120 DRMWRKTRSTMAGSSCLGVDPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALAD 178

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:| |:: ||||:::|||| :::||||     :::|||: :   | : :   |   |:||
Db        179 FIRNNLSTIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYT 237

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || |||||||
Db        238 YGPGATTIYPAAGGSDDWSYDQGIKYSFT 266


RESULT 13
AAR75132
ID   AAR75132 standard; Protein; 306 AA.
XX
AC   AAR75132;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine carboxypeptidase B.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB.
XX
OS   Sus scrofa.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   N-PSDB; AAQ90601.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Example 1; Page 18; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence   306 AA;

  Query Match             43.5%;  Score 582;  DB 16;  Length 306;
  Best Local Similarity   44.3%;  Pred. No. 3.2e-53;
  Matches  117;  Conservative   40;  Mismatches   67;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| : :|  :||:::: ||::|    :|:||| ||    | ||::|||
Db          6 YEKYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCG 64

Qy         66 IHAREWISPAFCLWFI-----------------------------GH------NRMWR 88
               ||||||| ||| ||:                             |:      |||||
Db         65 FHAREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWR 124

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  |   ||||| ||||    ||  |||:  | |||||   ||| | ||:| |:| |
Db        125 KTRSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNN 183

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::||||| |::|||      ::: ||: :|  ||: :  |   |:||:| |:
Db        184 LSSIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGA 242

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | || ||| || |||||||
Db        243 TTIYPAAGGSDDWAYDQGIKYSFT 266


RESULT 14
AAU04477
ID   AAU04477 standard; Protein; 306 AA.
XX
AC   AAU04477;
XX
DT   26-SEP-2001  (first entry)
XX
DE   Porcine carboxypeptidase B (CpB) protein.
XX
KW   Carboxypeptidase B; CpB; pig; protease activity; trypsin; chymotrypsin;
KW   animal product; purification; contaminant; biopharmaceutical agent;
KW   zinc-containing pancreatic exopeptidase; insulin; proteolytic processing;
KW   protein sequence analysis.
XX
OS   Sus scrofa.
XX
PN   WO200151624-A2.
XX
PD   19-JUL-2001.
XX
PF   12-JAN-2001; 2001WO-US00008.
XX
PR   12-JAN-2000; 2000US-0175781.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Hanquier JM;
XX
DR   WPI; 2001-442143/47.
XX
PT   New carboxypeptidase B, which is free of animal products and
PT   contaminating enzyme activity, useful for producing biopharmaceutical
PT   agents (e.g. insulin) and biologically active polypeptides, as well as
PT   in protein sequence analysis  -
XX
PS   Claim 4; Page 41-42; 42pp; English.
XX
CC   The sequence represents porcine carboxypeptidase B (CpB), which is
CC   essentially free of contaminating protease activity (e.g. trypsin or
CC   chymotrypsin activity) and free of animal products, bacterial cell
CC   components and other contaminants that result from purification
CC   processes. CpB is a zinc-containing pancreatic exopeptidase which
CC   specifically removes C-terminal arginine, lysine and ornithine from
CC   peptides. The CpB of the invention is useful for commercial and research
CC   purposes, e.g. in production of biopharmaceutical agents, such as insulin
CC   and other biologically active polypeptides, as well as in protein
CC   sequence analysis. Compositions containing CpB are useful in preparing
CC   proteins which are in need of proteolytic processing in order to form the
CC   mature protein.
XX
SQ   Sequence   306 AA;

  Query Match             43.5%;  Score 582;  DB 22;  Length 306;
  Best Local Similarity   44.3%;  Pred. No. 3.2e-53;
  Matches  117;  Conservative   40;  Mismatches   67;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| : :|  :||:::: ||::|    :|:||| ||    | ||::|||
```

```
Db          6 YEKYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCG 64

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              ||||||| ||| ||:                              |:      |||||
Db         65 FHAREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWR 124

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  |   ||||| ||||    ||  |||:  | |||||   ||| | ||:| |:| |
Db        125 KTRSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNN 183

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::||||| |::|||     ::: ||: :|  ||: :  |   |:||:| |:
Db        184 LSSIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGA 242

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | || ||| || |||||||
Db        243 TTIYPAAGGSDDWAYDQGIKYSFT 266


RESULT 15
AAR75131
ID   AAR75131 standard; Protein; 404 AA.
XX
AC   AAR75131;
XX
DT   13-MAR-1996  (first entry)
XX
DE   Porcine Tyr-His-Met Procarboxypeptidase B.
XX
KW   Procarboxypeptidase B;carboxypeptidase B; Pichia; pCPB;
KW   human serum albumin; premating factor alpha; mating factor alpha;
KW   proCBB.
XX
OS   Sus scrofa.
XX
PN   WO9514096-A1.
XX
PD   26-MAY-1995.
XX
PF   16-NOV-1994;   94WO-US13142.
XX
PR   16-NOV-1993;   93US-0153258.
XX
PA   (ELIL ) LILLY & CO ELI.
XX
PI   Fayerman JT,  Greenen DP,  Hershberger CL,  Larson JL;
PI   Sterner JL,  Zhang H;
XX
DR   WPI; 1995-200386/26.
DR   N-PSDB; AAQ90600.
XX
PT   DNA encoding porcine carboxy:peptidase B - used for transforming
PT   host cells, partic. Pichia species, for prodn. of the enzyme
XX
PS   Example 1; Page 16-17; 34pp; English.
XX
CC   The porcine carboxypeptidase B coding sequence can be place in a
CC   bacterial or pref. Pichia yeast expression vector. The expression
CC   vector further comprises the signal peptide of either human serum
CC   albumin (designated pLGD23 - NRRL B-21029); premating factor alpha
CC   (designated pFJ489 - NRRL B-21028); mating factor alpha (designated
CC   pFJ474 - NRRL B-21032) or the porcine proCBB signal peptide,
CC   (designated pLGD27 - NRRL B-21027).  The method can be used for
CC   producing large amounts of porcine carboxypeptidase B and when
CC   produced in Pichia yeast, the protein does not need solubilisation
CC   or folding.  The produced enzyme is then used for pref. cleaving
CC   basic residues from the carboxy terminus of proteins.
XX
SQ   Sequence   404 AA;

  Query Match            43.5%;  Score 582;  DB 16;  Length 404;
  Best Local Similarity  44.3%;  Pred. No. 4.9e-53;
  Matches  117;  Conservative   40;  Mismatches   67;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| : :|  :||:::: ||::|    :|:||| ||    | ||::|||
Db        104 YEKYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCG 162

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              ||||||| ||| ||:                              |:      |||||
Db        163 FHAREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWR 222

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  |   ||||| ||||    ||  |||:  | |||||   ||| | ||:| |:| |
Db        223 KTRSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNN 281

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::||||| |::|||     ::: ||: :|  ||: :  |   |:||:| |:
Db        282 LSSIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGA 340

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | || ||| || |||||||
Db        341 TTIYPAAGGSDDWAYDQGIKYSFT 364


Search completed: January  1, 2004, 19:28:08
Job time : 44.8831 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:21:46 ; Search time 15.1503 Seconds
(without alignments)
687.014 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1257	93.9	386	4	US-09-813-133A-2	Sequence 2, Appli
2	1239.5	92.6	423	1	US-07-649-591B-3	Sequence 3, Appli
3	1239.5	92.6	423	1	US-08-277-540-3	Sequence 3, Appli
4	1239.5	92.6	423	1	US-08-430-787A-3	Sequence 3, Appli
5	1239.5	92.6	423	2	US-08-869-057-2	Sequence 2, Appli
6	1228.5	91.8	423	4	US-09-813-133A-4	Sequence 4, Appli
7	610	45.6	307	2	US-08-782-760-6	Sequence 6, Appli
8	610	45.6	307	5	PCT-US96-00995-6	Sequence 6, Appli
9	582	43.5	306	1	US-08-696-139-4	Sequence 4, Appli
10	582	43.5	404	1	US-08-696-139-2	Sequence 2, Appli
11	567	42.4	329	4	US-09-011-769A-51	Sequence 51, Appl
12	567	42.4	349	4	US-09-011-769A-47	Sequence 47, Appl
13	567	42.4	415	2	US-08-860-882A-57	Sequence 57, Appl
14	567	42.4	415	4	US-09-011-769A-39	Sequence 39, Appl
15	567	42.4	424	4	US-09-011-769A-56	Sequence 56, Appl
16	561	41.9	417	1	US-07-649-591B-7	Sequence 7, Appli
17	561	41.9	417	1	US-08-277-540-7	Sequence 7, Appli
18	561	41.9	417	1	US-08-430-787A-7	Sequence 7, Appli
19	560	41.9	349	4	US-09-011-769A-60	Sequence 60, Appl
20	559	41.8	349	4	US-09-011-769A-64	Sequence 64, Appl
21	554	41.4	396	1	US-07-649-591B-4	Sequence 4, Appli
22	554	41.4	396	1	US-08-277-540-4	Sequence 4, Appli
23	554	41.4	396	1	US-08-430-787A-4	Sequence 4, Appli
24	552	41.3	417	1	US-07-649-591B-6	Sequence 6, Appli
25	552	41.3	417	1	US-08-277-540-6	Sequence 6, Appli
26	552	41.3	417	1	US-08-430-787A-6	Sequence 6, Appli
27	549	41.0	437	4	US-09-675-305-10	Sequence 10, Appl
28	549	41.0	613	3	US-09-171-945-113	Sequence 113, App
29	549	41.0	716	3	US-09-171-945-125	Sequence 125, App
30	492	36.8	399	4	US-09-710-099-8	Sequence 8, Appli
31	487.5	36.4	419	1	US-07-649-591B-5	Sequence 5, Appli
32	487.5	36.4	419	1	US-08-277-540-5	Sequence 5, Appli
33	487.5	36.4	419	1	US-08-430-787A-5	Sequence 5, Appli
34	483	36.1	417	1	US-07-649-591B-8	Sequence 8, Appli
35	483	36.1	417	1	US-08-277-540-8	Sequence 8, Appli
36	483	36.1	417	1	US-08-430-787A-8	Sequence 8, Appli
37	471.5	35.2	350	4	US-09-675-305-12	Sequence 12, Appl
38	465.5	34.8	419	3	US-08-640-906-2	Sequence 2, Appli
39	465.5	34.8	419	4	US-09-395-936-2	Sequence 2, Appli
40	465.5	34.8	436	4	US-09-710-099-6	Sequence 6, Appli
41	460.5	34.4	419	3	US-08-640-906-17	Sequence 17, Appl
42	460.5	34.4	419	4	US-09-395-936-17	Sequence 17, Appl
43	459	34.3	417	3	US-08-640-906-4	Sequence 4, Appli
44	459	34.3	417	4	US-09-395-936-4	Sequence 4, Appli
45	458	34.2	417	3	US-08-640-906-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-09-813-133A-2
; Sequence 2, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
; APPLICANT: GAN, Weiniu et al
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001173
; CURRENT APPLICATION NUMBER: US/09/813,133A
; CURRENT FILING DATE: 2001-06-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Human
US-09-813-133A-2

Query Match 93.9%; Score 1257; DB 4; Length 386;
Best Local Similarity 99.1%; Pred. No. 9.6e-129;
Matches 230; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 234

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 294

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        295 ELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 346
```

RESULT 2
US-07-649-591B-3
; Sequence 3, Application US/07649591B
; Patent No. 5206161
; GENERAL INFORMATION:
; APPLICANT: Dennis Drayna and Daniel Eaton
; TITLE OF INVENTION: No. 5206161el Plasma Carboxypeptidase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:

```
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/649,591B
;      FILING DATE:  19910201
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/266-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
US-07-649-591B-3

  Query Match             92.6%;  Score 1239.5;  DB 1;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 8.9e-127;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 3
US-08-277-540-3
; Sequence 3, Application US/08277540
; Patent No. 5474901
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5474901el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,540
;      FILING DATE:  19-JUL-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-277-540-3

  Query Match             92.6%;  Score 1239.5;  DB 1;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 8.9e-127;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 4
US-08-430-787A-3
; Sequence 3, Application US/08430787A
; Patent No. 5593674
;  GENERAL INFORMATION:
;    APPLICANT:  Drayna, Dennis T., Eaton, Dan L.
;    TITLE OF INVENTION:  No. 5593674el Plasma Carboxypeptidase
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/430,787A
;      FILING DATE:  27-APR-1995
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,540
;      FILING DATE:  19-JUL-1994
;      APPLICATION NUMBER:  08/167727
;      FILING DATE:  15-DEC-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/959944
;      FILING DATE:  14-OCT-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/649591
;      FILING DATE:  01-FEB-91
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hasak, Janet E.
;      REGISTRATION NUMBER:  28,616
;      REFERENCE/DOCKET NUMBER:  689D1C1D1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-1896
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
US-08-430-787A-3

  Query Match             92.6%;  Score 1239.5;  DB 1;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 8.9e-127;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 5
US-08-869-057-2
; Sequence 2, Application US/08869057
; Patent No. 5985562
;  GENERAL INFORMATION:
;    APPLICANT:  Morser, Michael J
;    APPLICANT:  Nagashima, Mariko
;    TITLE OF INVENTION:  Method of Detecting Thrombotic Disease
;    TITLE OF INVENTION:  Risk
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Berlex Biosciences Legal Department
;      STREET:  15049 San Pablo Avenue
;      CITY:  Richmond
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94804-0099
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/869,057
;      FILING DATE:  03-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Washtien, Wendy L
;      REGISTRATION NUMBER:  36,301
;      REFERENCE/DOCKET NUMBER:  51509AUSM1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  510-262-5411
;      TELEFAX:  510-262-7095
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  423 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      TISSUE TYPE:  Plasma
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  23..401
US-08-869-057-2

  Query Match             92.6%;  Score 1239.5;  DB 2;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 8.9e-127;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 6
US-09-813-133A-4
; Sequence 4, Application US/09813133A
; Patent No. 6455294
; GENERAL INFORMATION:
;  APPLICANT: GAN, Weiniu et al
;  TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
;  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
;  TITLE OF INVENTION:  USES THEREOF
;  FILE REFERENCE: CL001173
;  CURRENT APPLICATION NUMBER: US/09/813,133A
;  CURRENT FILING DATE:  2001-06-06
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
```

```
; SEQ ID NO 4
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Human
US-09-813-133A-4

  Query Match            91.8%;  Score 1228.5;  DB 4;  Length 423;
  Best Local Similarity  85.5%;  Pred. No. 1.4e-125;
  Matches  230;  Conservative    0;  Mismatches    2;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 7
US-08-782-760-6
; Sequence 6, Application US/08782760
; Patent No. 5948668
;  GENERAL INFORMATION:
;    APPLICANT:  Hartman, Jacob
;    APPLICANT:  Fulga, Netta
;    APPLICANT:  Mendelovitch, Simona
;    APPLICANT:  Gorecki, Marian
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/782,760
;      FILING DATE:  13-JAN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/378,233
;      FILING DATE:  25-JAN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  307 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-782-760-6

  Query Match            45.6%;  Score 610;  DB 2;  Length 307;
  Best Local Similarity  45.0%;  Pred. No. 2.8e-58;
  Matches  121;  Conservative   43;  Mismatches   65;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||   | :|::   | :||: :   :||::|:  ||::||   :|||| || :  | ||
Db          1 ASGHSYTKYNNWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAI 59

Qy         61 WIDCGIHAREWISPAFCLWFI-------------------------------GH------ 83
              :|||| ||||||||||| ||:                               |:
Db         60 FIDCGFHAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTK 119

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :||||| ||  | : |:| | |||| :  ||| ||| | |||||||  |||| | ||:|
Db        120 DRMWRKTRSTMAGSSCLGVDPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALAD 178

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:| |:: ||||:::|||||  :::||||     :::|||: :   | : :  |   |:||
Db        179 FIRNNLSTIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYT 237

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || |||||||
Db        238 YGPGATTIYPAAGGSDDWSYDQGIKYSFT 266


RESULT 8
PCT-US96-00995-6
; Sequence 6, Application PC/TUS9600995
;  GENERAL INFORMATION:
;    APPLICANT:  Bio-Technology General Corp.
;    TITLE OF INVENTION:  PRODUCTION OF ENZYMATICALLY ACTIVE
;    TITLE OF INVENTION:  CARBOXYPEPTIDASE B
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham LLP
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  U.S.A.
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US96/00995
;      FILING DATE:  25-JAN-1996
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  0336/43847-A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 278-0400
;      TELEFAX:  (212) 391-0525
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;     LENGTH: 308 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US96-00995-6

  Query Match            45.6%;  Score 610;  DB 5;  Length 307;
  Best Local Similarity  45.0%;  Pred. No. 2.8e-58;
  Matches  121;  Conservative   43;  Mismatches   65;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
```

```
Db          1 ASGHSYTKYNNWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAI 59

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
Db         60 FIDCGFHAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTK 119

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
Db        120 DRMWRKTRSTMAGSSCLGVDPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALAD 178

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
Db        179 FIRNNLSTIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYT 237

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
Db        238 YGPGATTIYPAAGGSDDWSYDQGIKYSFT 266
```

RESULT 9
US-08-696-139-4
; Sequence 4, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hershberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-139-4

Query Match 43.5%; Score 582; DB 1; Length 306;
Best Local Similarity 44.3%; Pred. No. 3.2e-55;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

```
Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
Db          6 YEKYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCG 64

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
Db         65 FHAREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWR 124

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
Db        125 KTRSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNN 183

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
Db        184 LSSIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGA 242

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
Db        243 TTIYPAAGGSDDWAYDQGIKYSFT 266
```

RESULT 10
US-08-696-139-2
; Sequence 2, Application US/08696139
; Patent No. 5672496
; GENERAL INFORMATION:
; APPLICANT: Fayerman, Jeffrey T.
; APPLICANT: Greenen, David P.
; APPLICANT: Hershberger, Charles L.
; APPLICANT: Larson, Jeffrey L.
; APPLICANT: Sterner, Jane L.
; APPLICANT: Zhang, Haichao
; TITLE OF INVENTION: DNA SEQUENCES ENCODING PORCINE
; TITLE OF INVENTION: PANCREATIC CARBOXYPEPTIDASE B
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/696,139
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,258
; FILING DATE: 16-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-8681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-696-139-2

Query Match 43.5%; Score 582; DB 1; Length 404;
Best Local Similarity 44.3%; Pred. No. 4.8e-55;
Matches 117; Conservative 40; Mismatches 67; Indels 40; Gaps 5;

```
Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
```

```
                ||:|::   | :| : :|  :||:::: ||::|    :|:||| ||    | ||::|||
Db        104 YEKYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCG 162

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
               ||||||| ||| ||:                              |:      |||||
Db        163 FHAREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWR 222

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
               | ||  |   ||||| ||||    ||  |||:  | |||||   ||| | ||:| |:| |
Db        223 KTRSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNN 281

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
               :: ||||:::||||| |::||||     ::: ||: :|  ||: :  |   |:||:| |:
Db        282 LSSIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGA 340

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
               |:| | || ||| || |||||||
Db        341 TTIYPAAGGSDDWAYDQGIKYSFT 364


RESULT 11
US-09-011-769A-51
; Sequence 51, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 51:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 329 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-011-769A-51

  Query Match           42.4%;  Score 567;  DB 4;  Length 329;
  Best Local Similarity  41.3%;  Pred. No. 1.5e-53;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
```

```
                |:   ||:|:    | :| : :   :| :::: ||::||   :|:||| ||    | ||
Db         23 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 81

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
               ::||| |||||||||| ||:                              |:
Db         82 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 141

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
               :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        142 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 200

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
               |:|  :: ||||:::||||| :::|||     ::: ||: :|   |: :  :   |:||
Db        201 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 259

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
               :| |: |:| | || ||| || ||:||||
Db        260 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 288


RESULT 12
US-09-011-769A-47
; Sequence 47, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 47:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 349 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 47:
US-09-011-769A-47

  Query Match           42.4%;  Score 567;  DB 4;  Length 349;
  Best Local Similarity  41.3%;  Pred. No. 1.7e-53;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
```

```
                |:  ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db         23 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 81

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
               ::||| |||||||||||| ||:                              |:
Db         82 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 141

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
               :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        142 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 200

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
               |:|  :: ||||:::||||| :::|||| :::  ||: :|  |: :  :  |:||
Db        201 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 259

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
               :| |: |:| | || ||| || ||:||||
Db        260 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 288


RESULT 13
US-08-860-882A-57
; Sequence 57, Application US/08860882A
; Patent No. 5985281
;  GENERAL INFORMATION:
;    APPLICANT:  TAYLORSON, CHRISTOPHER JOHN
;    APPLICANT:  EGGELTE, HENDRIKUS JOHANNES
;    APPLICANT:  TARRAGONA-FIOL, ANTONIO
;    APPLICANT:  RABIN, BRIAN ROBERT
;    APPLICANT:  BOYLE, FRANCIS THOMAS
;    APPLICANT:  HENNAM,JOHN FREDERICK
;    APPLICANT:  BLAKELY, DAVID CHARLES
;    APPLICANT:  MARSHAM, PETER ROBERT
;    APPLICANT:  HEATON, DAVID WILLIAM
;    APPLICANT:  DAVIES, DAVID HUW
;    TITLE OF INVENTION:  CHEMICAL COMPOUNDS
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  PILLSBURY, MADISON & SUTRO
;      STREET:  1100 NEW YORK AVENUE, N.W.
;      CITY:  WASHINGTON
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy Disk
;      COMPUTER:  IBM compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/860,882A
;      FILING DATE:  JUNE 23, 1997
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DONALD J. BIRD
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  9901/238653
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3027
;      TELEFAX:  (202) 822-0944
;      TELEX:  6174627 CUSH
;  INFORMATION FOR SEQ ID NO:  57:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  415 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-860-882A-57

  Query Match           42.4%;  Score 567;  DB 2;  Length 415;
  Best Local Similarity   41.3%;  Pred. No. 2.2e-53;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;
```

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
               |:   ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db        109 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 167

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
               ::||| |||||||||||| ||:                              |:
Db        168 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 227

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
               :| |||| ||  :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        228 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 286

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
               |:|  :: ||||:::||||| :::||||     ::: ||: :|   |: :  :  |:||
Db        287 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 345

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
               :| |: |:| | || ||| || ||:||||
Db        346 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 374


RESULT 14
US-09-011-769A-39
; Sequence 39, Application US/09011769A
; Patent No. 6436691
;  GENERAL INFORMATION:
;       APPLICANT: SLATER, Anthony M.
;                  BLAKEY, David C.
;                  DAVIES, David H.
;                  HENNAM, John F.
;                  HENNEQUIN, Laurent F.A.
;                  MARSHAM, Peter R.
;                  DOWELL, Robert I.
;       TITLE OF INVENTION: Chemical Compounds
;       NUMBER OF SEQUENCES: 87
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Pillsbury Madison & Sutro, LLP
;            STREET: 1100 New York Ave., N.W.
;            CITY: Washington
;            STATE: D.C.
;            COUNTRY: U.S.A.
;            ZIP: 20005
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: 1.44 Mb disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: MS Word
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/011,769A
;            FILING DATE: 13-Feb-1998
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: PCT/GB96/01975
;            FILING DATE: 13-AUG-1996
;            APPLICATION NUMBER: GB 9612295.7
;            FILING DATE: 12-JUN-1996
;            APPLICATION NUMBER: GB 9611019.2
;            FILING DATE: 25-MAY-1996
;            APPLICATION NUMBER: GB 9516810.0
;            FILING DATE: 16-AUG-1995
;  INFORMATION FOR SEQ ID NO: 39:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 415 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: peptide
;       SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-011-769A-39

  Query Match           42.4%;  Score 567;  DB 4;  Length 415;
  Best Local Similarity   41.3%;  Pred. No. 2.2e-53;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;
```

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db        109 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 167

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              ::||| ||||||||||| ||:                              |:
Db        168 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 227

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        228 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 286

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::||||| :::||||     ::: ||: :|   |: : :   |:||
Db        287 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 345

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || ||:||||
Db        346 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 374
```

```
RESULT 15
US-09-011-769A-56
; Sequence 56, Application US/09011769A
; Patent No. 6436691
;   GENERAL INFORMATION:
;        APPLICANT: SLATER, Anthony M.
;                   BLAKEY, David C.
;                   DAVIES, David H.
;                   HENNAM, John F.
;                   HENNEQUIN, Laurent F.A.
;                   MARSHAM, Peter R.
;                   DOWELL, Robert I.
;        TITLE OF INVENTION: Chemical Compounds
;        NUMBER OF SEQUENCES: 87
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Pillsbury Madison & Sutro, LLP
;             STREET: 1100 New York Ave., N.W.
;             CITY: Washington
;             STATE: D.C.
;             COUNTRY: U.S.A.
;             ZIP: 20005
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 1.44 Mb disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: MS Word
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/011,769A
;             FILING DATE: 13-Feb-1998
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: PCT/GB96/01975
;             FILING DATE: 13-AUG-1996
;             APPLICATION NUMBER: GB 9612295.7
;             FILING DATE: 12-JUN-1996
;             APPLICATION NUMBER: GB 9611019.2
;             FILING DATE: 25-MAY-1996
;             APPLICATION NUMBER: GB 9516810.0
;             FILING DATE: 16-AUG-1995
;   INFORMATION FOR SEQ ID NO: 56:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 424 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 56:
US-09-011-769A-56

  Query Match             42.4%;  Score 567;  DB 4;  Length 424;
  Best Local Similarity   41.3%;  Pred. No. 2.3e-53;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;
```

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db        118 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 176

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              ::||| ||||||||||| ||:                              |:
Db        177 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 236

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        237 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 295

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::||||| :::||||     ::: ||: :|   |: : :   |:||
Db        296 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || ||:||||
Db        355 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 383
```

```
Search completed: January  1, 2004, 19:33:11
Job time : 16.1503 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:32:17 ; Search time 35.1795 Seconds
(without alignments)
1403.024 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 733937 seqs, 200641211 residues

Total number of hits satisfying chosen parameters: 733937

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1239.5	92.6	423	12	US-10-379-836-17	Sequence 17, Appl
2	1164.5	87.0	423	12	US-10-379-836-2	Sequence 2, Appli
3	1105.5	82.6	422	12	US-10-379-836-18	Sequence 18, Appl
4	1095.5	81.9	422	12	US-10-379-836-16	Sequence 16, Appl
5	816.5	61.0	211	9	US-09-925-302-467	Sequence 467, App
6	567	42.4	402	12	US-10-379-836-20	Sequence 20, Appl
7	552	41.3	417	12	US-10-341-434-188	Sequence 188, App
8	552	41.3	417	15	US-10-229-546-2	Sequence 2, Appli
9	552	41.3	417	15	US-10-229-546-9	Sequence 9, Appli
10	550.5	41.1	374	9	US-09-888-615-61	Sequence 61, Appl
11	550.5	41.1	444	16	US-10-176-306-74	Sequence 74, Appl
12	549	41.0	437	12	US-10-274-639-12	Sequence 12, Appl
13	549	41.0	437	14	US-10-200-344-10	Sequence 10, Appl
14	549	41.0	613	10	US-09-910-059-113	Sequence 113, App
15	549	41.0	716	10	US-09-910-059-125	Sequence 125, App
16	501.5	37.5	298	12	US-10-379-836-19	Sequence 19, Appl
17	492	36.8	399	15	US-10-200-910-8	Sequence 8, Appli
18	471.5	35.2	350	14	US-10-200-344-12	Sequence 12, Appl
19	465.5	34.8	406	9	US-09-925-297-517	Sequence 517, App
20	465.5	34.8	419	12	US-10-345-680-65	Sequence 65, Appl
21	465.5	34.8	436	15	US-10-200-910-6	Sequence 6, Appli
22	458	34.2	421	11	US-09-946-374-234	Sequence 234, App
23	458	34.2	421	12	US-10-015-387A-234	Sequence 234, App
24	458	34.2	421	12	US-10-006-130A-234	Sequence 234, App
25	458	34.2	421	12	US-10-199-672-308	Sequence 308, App
26	458	34.2	421	12	US-10-006-172A-234	Sequence 234, App
27	458	34.2	421	12	US-10-187-749-308	Sequence 308, App
28	458	34.2	421	12	US-10-194-457-308	Sequence 308, App
29	458	34.2	421	12	US-10-184-642-308	Sequence 308, App
30	458	34.2	421	12	US-10-196-747-308	Sequence 308, App
31	458	34.2	421	12	US-10-015-392A-234	Sequence 234, App
32	458	34.2	421	12	US-10-017-253A-234	Sequence 234, App
33	458	34.2	421	12	US-10-173-689-308	Sequence 308, App
34	458	34.2	421	12	US-10-173-690-308	Sequence 308, App
35	458	34.2	421	12	US-10-173-691-308	Sequence 308, App
36	458	34.2	421	12	US-10-173-692-308	Sequence 308, App
37	458	34.2	421	12	US-10-173-694-308	Sequence 308, App
38	458	34.2	421	12	US-10-173-698-308	Sequence 308, App
39	458	34.2	421	12	US-10-173-699-308	Sequence 308, App
40	458	34.2	421	12	US-10-173-707-308	Sequence 308, App
41	458	34.2	421	12	US-10-174-569-308	Sequence 308, App
42	458	34.2	421	12	US-10-174-583-308	Sequence 308, App
43	458	34.2	421	12	US-10-174-587-308	Sequence 308, App
44	458	34.2	421	12	US-10-174-589-308	Sequence 308, App
45	458	34.2	421	12	US-10-174-591-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-10-379-836-17
; Sequence 17, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
; TITLE OF INVENTION: TAFI
; FILE REFERENCE: D0214NP
; CURRENT APPLICATION NUMBER: US/10/379,836
; CURRENT FILING DATE: 2003-03-04
; PRIOR APPLICATION NUMBER: U.S. 60/361,523
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 423
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-836-17

Query Match 92.6%; Score 1239.5; DB 12; Length 423;
Best Local Similarity 86.2%; Pred. No. 2.5e-121;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              ||||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 2
US-10-379-836-2
; Sequence 2, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
;   LENGTH: 423
;   TYPE: PRT
;   ORGANISM: Papio hamadryas
US-10-379-836-2

  Query Match             87.0%;  Score 1164.5;  DB 12;  Length 423;
  Best Local Similarity   80.7%;  Pred. No. 1.8e-113;
  Matches  217;  Conservative    7;  Mismatches    8;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||| |||::||||||||||||:||:||||||||||||||||:
Db        115 ASASYYEQYHSLNEIYSWIELITEKYPDMLTKIHIGSSYEKHPLYVLKVSGKEQTAKNAM 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITEYYGIIGEYTNLLRHVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||| |||||||||||||||||||| |||||||||||||||| ||||:
Db        235 NRMWRKNRSFYANNRCIGTDLNRNFASKHWCEEGASSFSCSETYCGLYPESEPEAKAVAN 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||| |||||||||||||||||||||||||||||||||||||||||:||||| |||
Db        295 FLRRNINHIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIQKTSKNIRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              || |||||||||| ||||||||||||||
Db        355 HGRGSETLYLAPGGADDWIYDLGIKYSFT 383


RESULT 3
US-10-379-836-18
; Sequence 18, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-10-379-836-18

  Query Match             82.6%;  Score 1105.5;  DB 12;  Length 422;
  Best Local Similarity   77.0%;  Pred. No. 2.8e-107;
  Matches  207;  Conservative   10;  Mismatches   15;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||| |||:||||| ||:|||||||||||||||||  ||||
Db        114 ASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAI 173

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||:
Db        174 WIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYIMPVMNVDGYDYTWKK 233

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||| : || |:|||||||||||||:|||||||||||||||||||||||||||||
Db        234 NRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSETYCGLYPESEPEVKAVAD 293

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||: |||||||||||| |:||||| |||||||||||||||||||||  :||||||
Db        294 FLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYT 353

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              || |||:||||||| ||||||||||||||
Db        354 HGSGSESLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 4
US-10-379-836-16
; Sequence 16, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAFI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16
;   LENGTH: 422
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-10-379-836-16

  Query Match             81.9%;  Score 1095.5;  DB 12;  Length 422;
  Best Local Similarity   75.8%;  Pred. No. 3.2e-106;
  Matches  204;  Conservative   11;  Mismatches   17;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||:|||||||||||||||| |||:||||| ||:||||:||||||||||||  ||||
Db        114 ASSSYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSYEKYPLYVLKVSGKEHRVKNAI 173

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||:
Db        174 WIDCGIHAREWISPAFCLWFIGYVTQFHGKENTYTRLLRHVDFYIMPVMNVDGYDYTWKK 233

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||| : || |:|||||||||||||:|||| ||||||||||||||||||||||||
Db        234 NRMWRKNRSVHMNNRCVGTDLNRNFASKHWCEKGASSFSCSETYCGLYPESEPEVKAVAD 293

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||| |||||||||||| |:||||| |||||||||||||||||||||  :||||||
Db        294 FLRRNININHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYT 353

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              || |||:||||||| ||||||||||||||
Db        354 HGSGSESLYLAPGGSDDWIYDLGIKYSFT 382
```

```
RESULT 5
US-09-925-302-467
; Sequence 467, Application US/09925302
; Patent No. US20020044941A1
; GENERAL INFORMATION:
;  APPLICANT: Rosen et al.
;  TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;  FILE REFERENCE: PA104
;  CURRENT APPLICATION NUMBER: US/09/925,302
;  CURRENT FILING DATE:  2001-08-10
;  PRIOR APPLICATION NUMBER: PCT/US00/05918
;  PRIOR FILING DATE: 2000-03-08
;  PRIOR APPLICATION NUMBER: 60/124,270
;  PRIOR FILING DATE: 1999-03-12
;  NUMBER OF SEQ ID NOS: 896
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 467
;   LENGTH: 211
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-925-302-467

  Query Match             61.0%;  Score 816.5;  DB 9;  Length 211;
  Best Local Similarity   90.6%;  Pred. No. 2.3e-77;
  Matches  154;  Conservative    1;  Mismatches    6;  Indels    9;  Gaps    2;

Qy         71 WISPAFCLWFI--------GHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 122
              |||   | | :          ||||||||||||||||||||||||||||||||||||||
Db          3 WIS-MLCRWLMWMVMNYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSS 61

Qy        123 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 182
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 CSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEEL 121

Qy        183 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        122 SLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFT 171


RESULT 6
US-10-379-836-20
; Sequence 20, Application US/10379836
; Publication No. US20030215850A1
; GENERAL INFORMATION:
;  APPLICANT: Bristol-Myers Squibb Company
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND POLYPEPTIDES ENCODING BABOON
;  TITLE OF INVENTION:  TAPI
;  FILE REFERENCE: D0214NP
;  CURRENT APPLICATION NUMBER: US/10/379,836
;  CURRENT FILING DATE:  2003-03-04
;  PRIOR APPLICATION NUMBER: U.S. 60/361,523
;  PRIOR FILING DATE: 2002-03-04
;  NUMBER OF SEQ ID NOS: 24
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
;   LENGTH: 402
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-379-836-20

  Query Match             42.4%;  Score 567;  DB 12;  Length 402;
  Best Local Similarity   41.3%;  Pred. No. 8.6e-51;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db         96 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 154

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              ::||| |||||||||| ||:                              |:
Db        155 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLNKLDFYVLPVLNIDGYIYTWTK 214

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        215 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 273

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::||||| :::|||     ::: ||: :|   |: :  :   |:||
Db        274 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 332

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:|.| || ||| || ||:||||
Db        333 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 361


RESULT 7
US-10-341-434-188
; Sequence 188, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
;  APPLICANT: OriGene Technologies
;  TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
;  FILE REFERENCE: 9U 204 205 R1
;  CURRENT APPLICATION NUMBER: US/10/341,434
;  CURRENT FILING DATE:  2003-07-18
;  PRIOR APPLICATION NUMBER: US 60/348,164
;  PRIOR FILING DATE: 2002-01-15
;  PRIOR APPLICATION NUMBER: US 60/348,119
;  PRIOR FILING DATE: 2002-01-15
;  NUMBER OF SEQ ID NOS: 238
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 188
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-341-434-188

  Query Match             41.3%;  Score 552;  DB 12;  Length 417;
  Best Local Similarity   41.4%;  Pred. No. 3.4e-49;
  Matches  109;  Conservative   45;  Mismatches   69;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|::  :| :|  | : :::|:|:::| |||: |  ||||||: |::   : ||::|||
Db        116 YAKYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              ||||||:||||| ||:                              |:      |||||
Db        175 IHAREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   |: |||||||||| :  |      ::  |:: | |  |||| | ||| :|:| :
Db        235 KNRSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :|:|| ||: ||||| ::||| ||      :||:|: ||      :  |   ||| :|
Db        294 LNEIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIE 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              |:|   |   || ||||||::|
Db        353 STIYPISGSSLDWAYDLGIKHTF 375


RESULT 8
US-10-229-546-2
; Sequence 2, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
```

```
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-2

  Query Match            41.3%;  Score 552;  DB 15;  Length 417;
  Best Local Similarity  41.4%;  Pred. No. 3.4e-49;
  Matches  109;  Conservative   45;  Mismatches   69;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||
Db        116 YAKYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              ||||||:|||||  ||:                              |:      |||||
Db        175 IHAREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   |: |||||||||| :  |      ::  |:: | |  |||| | ||| :|:| :
Db        235 KNRSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :|:|| ||: ||||| ::||| ||     :||:|: ||       :  |   ||| :|
Db        294 LNEIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIE 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              |:|   |   || ||||||::|
Db        353 STIYPISGSSLDWAYDLGIKHTF 375


RESULT 9
US-10-229-546-9
; Sequence 9, Application US/10229546
; Publication No. US20030082649A1
; GENERAL INFORMATION:
;  APPLICANT: Weich, Nadine S.
;  APPLICANT:  Carroll, Joseph M.
;  APPLICANT:  Millennium Pharmaceuticals, Inc.
;  TITLE OF INVENTION: 6299, A HUMAN ZINC CARBOXYPEPTIDASE
;  TITLE OF INVENTION:  FAMILY MEMBER AND USES THEREFOR
;  FILE REFERENCE: MPI01-156P1RNM
;  CURRENT APPLICATION NUMBER: US/10/229,546
;  CURRENT FILING DATE:  2002-08-28
;  PRIOR APPLICATION NUMBER: 60/316,575
;  PRIOR FILING DATE: 2001-08-31
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 417
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-229-546-9

  Query Match            41.3%;  Score 552;  DB 15;  Length 417;
  Best Local Similarity  41.4%;  Pred. No. 3.4e-49;
  Matches  109;  Conservative   45;  Mismatches   69;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||
Db        116 YAKYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              ||||||:|||||  ||:                              |:      |||||
Db        175 IHAREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   |: |||||||||| :  |      ::  |:: | |  |||| | ||| :|:| :
Db        235 KNRSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :|:|| ||: ||||| ::||| ||     :||:|: ||       :  |   ||| :|
Db        294 LNEIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIE 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              |:|   |   || ||||||::|
Db        353 STIYPISGSSLDWAYDLGIKHTF 375


RESULT 10
US-09-888-615-61
; Sequence 61, Application US/09888615
; Patent No. US20020064856A1
; GENERAL INFORMATION:
;  APPLICANT: PLOWMAN, GREGORY
;  APPLICANT:  WHYTE, DAVID
;  APPLICANT:  CAENEPEEL, SEAN
;  APPLICANT:  CHARYDCZAK, GLEN
;  APPLICANT:  MANNING, GERARD
;  APPLICANT:  SUDARSANAM, SUCHA
;  TITLE OF INVENTION: NOVEL PROTEASES
;  FILE REFERENCE: 038602/1214
;  CURRENT APPLICATION NUMBER: US/09/888,615
;  CURRENT FILING DATE:  2001-06-26
;  PRIOR APPLICATION NUMBER: 60/214,047
;  PRIOR FILING DATE: 2000-06-26
;  NUMBER OF SEQ ID NOS: 150
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 61
;   LENGTH: 374
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-888-615-61

  Query Match            41.1%;  Score 550.5;  DB 9;  Length 374;
  Best Local Similarity  39.4%;  Pred. No. 4.2e-49;
  Matches  104;  Conservative   45;  Mismatches   76;  Indels   39;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  || : ||| |:  |:|:: :::|:  :| ::| :|:| ||:|      |  ||:|||
Db         47 YNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCG 106

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              |||||||:|||| ||:                              |:      :|:||
Db        107 IHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWR 166

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              |:|| : |  | ||||||| :  ||  ||| :   :|:||  | |||| ||||||:
Db        167 KSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIESK 225

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
               : |  :::|||| | |: || ||::|| :| |:  |  :|  |: |    | |  |  :
Db        226 KDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANAL-KAKYGTNYRVGSSA 284

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              : || : |   ||  |:|| :|:|
Db        285 DILYASSGSSRDWARDIGIPFSYT 308


RESULT 11
US-10-176-306-74
; Sequence 74, Application US/10176306
; Publication No. US20030130485A1
; GENERAL INFORMATION:
;  APPLICANT: Meyers, Rachel A.
;  APPLICANT:  Curtis, Rory A. J.
;  APPLICANT:  Glücksmann, Maria Alexandra
;  APPLICANT:  Bandaru, Rajasekhar
;  APPLICANT:  Kapeller-Libermann, Rosana
;  TITLE OF INVENTION: NOVEL HUMAN GENES AND METHODS OF USE THEREOF
;  FILE REFERENCE: 10448-195001
;  CURRENT APPLICATION NUMBER: US/10/176,306
;  CURRENT FILING DATE:  2002-06-20
```

```
;  PRIOR APPLICATION NUMBER: 10/001,137
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: PCT/US01/45291
;  PRIOR FILING DATE: 2001-11-14
;  PRIOR APPLICATION NUMBER: 60/248,362
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,331
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/248,365
;  PRIOR FILING DATE: 2000-11-14
;  PRIOR APPLICATION NUMBER: 60/250,077
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,327
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 60/250,176
;  PRIOR FILING DATE: 2000-11-30
;  PRIOR APPLICATION NUMBER: 10/023,617
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: PCT/US01/49416
;  PRIOR FILING DATE: 2001-12-18
;  PRIOR APPLICATION NUMBER: 60/256,249
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 60/256,405
;  PRIOR FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: 10/083,248
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: PCT/US01/46717
;  PRIOR FILING DATE: 2001-10-22
;  PRIOR APPLICATION NUMBER: 60/242,324
;  PRIOR FILING DATE: 2000-10-20
;  PRIOR APPLICATION NUMBER: 60/242,518
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/241,989
;  PRIOR FILING DATE: 2000-10-20
;  NUMBER OF SEQ ID NOS: 86
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 74
;   LENGTH: 444
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-176-306-74

  Query Match             41.1%;  Score 550.5;  DB 16;  Length 444;
  Best Local Similarity   39.4%;  Pred. No. 5.3e-49;
  Matches  104;  Conservative   45;  Mismatches   76;  Indels   39;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  || : ||| |:  |:|:: :::|:  :| ::| :|:| ||:|      |  ||:|||
Db        177 YNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPIYYLKISQPSGNPKKIIWMDCG 236

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              |||||||:|||| ||:                              |:      :|:||
Db        237 IHAREWIAPAFCQWFVKEILQNHKDNSRIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWR 296

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              |:|| : |  | |||||||| :  ||  ||| :   :|:||  | |||| ||||||:
Db        297 KSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIESK 355

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
               : |  :::|||| | |: || ||::|| :| |:  |  :|  |: |    | |  |  :
Db        356 KDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANAL-KAKYGTNYRVGSSA 414

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              : || : |   ||  |:|| :|:|
Db        415 DILYASSGSSRDWARDIGIPFSYT 438


RESULT 12
US-10-274-639-12
; Sequence 12, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  DELEGEANE, Angelo M.; GANDHI, Ameena R.
;  APPLICANT:  HAFALIA, April J.A.; LU, Dyung Aina M.
;  APPLICANT:  PATTERSON, Chandra; TRIBOULEY, Catherine M.
;  APPLICANT:  DAS, Debopriya; KALLICK, Deborah A.
;  APPLICANT:  NGUYEN, Danniel B.; LEE, Ernestine A.
;  APPLICANT:  KHAN, Farrah A.; YUE, Henry
;  APPLICANT:  AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
;  APPLICANT:  POLICKY, Jennifer L.; RAMKUMAR, Jayalaxmi
;  APPLICANT:  YANG, Junming; THANGAVELU, Kavitha
;  APPLICANT:  DING, Li; KEARNEY, Liam
;  APPLICANT:  BAUGHN, Mariah R.; BOROWSKY, Mark L.
;  APPLICANT:  SANJANWALA, Madhusudan M.; YAO, Monique G.
;  APPLICANT:  BURFORD, Neil; WALIA, Narinder K.
;  APPLICANT:  LAL, Preeti G.; LEE, Sally
;  APPLICANT:  TODD, Stephen; LO, Terence P.
;  APPLICANT:  TANG, Y. Tom; ELLIOTT, Vicki S.
;  APPLICANT:  AZIMZAI, Yalda; LU, Yan
;  TITLE OF INVENTION: PROTEASES
;  FILE REFERENCE: PI-0167 USA
;  CURRENT APPLICATION NUMBER: US/10/274,639
;  CURRENT FILING DATE:  2002-10-18
;  PRIOR APPLICATION NUMBER: PCT/US01/22397
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/220,063
;  PRIOR FILING DATE: 2000-07-21
;  PRIOR APPLICATION NUMBER: US 60/221,680
;  PRIOR FILING DATE: 2000-07-28
;  PRIOR APPLICATION NUMBER: US 60/223,544
;  PRIOR FILING DATE: 2000-08-04
;  PRIOR APPLICATION NUMBER: US 60/224,717
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/225,988
;  PRIOR FILING DATE: 2000-08-16
;  PRIOR APPLICATION NUMBER: US 60/227,568
;  PRIOR FILING DATE: 2000-08-23
;  NUMBER OF SEQ ID NOS: 42
;  SOFTWARE: PERL Program
; SEQ ID NO 12
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030232349A1 4948403CD1
US-10-274-639-12

  Query Match             41.0%;  Score 549;  DB 12;  Length 437;
  Best Local Similarity   40.3%;  Pred. No. 7.5e-49;
  Matches  106;  Conservative   40;  Mismatches   77;  Indels   40;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              || |||| || :|:  : : |  ::    || |:|   |::||: |:    | |:|||||
Db        136 YEVYHSLEEIQNWMHHLNKTHSGLIHMFSIGRSYEGRSLFILKL-GRRSRLKRAVWIDCG 194

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||||||| |||| ||:                                   ::| ||
Db        195 IHAREWIGPAFCQWFVKEALLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSWTNDRFWR 254

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  :   | | | |||:  | ||:||||   | :|||| :||||||||||:|||::
Db        255 KTRSRNSRFRCRGVDANRNWKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAVANFLRKH 313

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
               |:||:| |:|:| :::|||| :  :  :   | :|| |::      || :|  |
Db        314 RKHIRAYLSFHAYAQMLLYPYSYKYATIPNFRCVESAAYKAVNALQSV-YGVRYRYGPAS 372

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              |||:: |   || |  || |:|
Db        373 TTLYVSSGSSMDWAYKNGIPYAF 395


RESULT 13
US-10-200-344-10
```

```
; Sequence 10, Application US/10200344
; Publication No. US20020173641A1
; GENERAL INFORMATION:
;  APPLICANT: Donoho, Gregory
;  APPLICANT:  Turner, C. Alexander Jr.
;  APPLICANT:  Nehls, Michael
;  APPLICANT:  Friedrich, Glenn
;  APPLICANT:  Zambrowicz, Brian
;  APPLICANT:  Sands, Arthur T.
;  TITLE OF INVENTION: No. US20020173641A1el Human Carboxypeptidases and
;  TITLE OF INVENTION:  Polynucleotides Encoding the Same
;  FILE REFERENCE: LEX-0047-USA
;  CURRENT APPLICATION NUMBER: US/10/200,344
;  CURRENT FILING DATE:  2002-07-19
;  PRIOR APPLICATION NUMBER: US/09/675,305
;  PRIOR FILING DATE: 2000-09-29
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10
;   LENGTH: 437
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-10-200-344-10

  Query Match             41.0%;  Score 549;  DB 14;  Length 437;
  Best Local Similarity   40.3%;  Pred. No. 7.5e-49;
  Matches  106;  Conservative   40;  Mismatches   77;  Indels   40;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              || |||| || :|:  : : |  ::    || |:|   |::||: |:    | |:|||||
Db        136 YEVYHSLEEIQNWMHHLNKTHSGLIHMFSIGRSYEGRSLPILKL-GRRSRLKRAVWIDCG 194

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||||||| |||| ||:                                   ::| ||
Db        195 IHAREWIGPAFCQWFVKEALLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSWTNDRFWR 254

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  :   | | | |||:  | ||:||||   | :|||| :||||||||||:|||::
Db        255 KTRSRNSRFRCRGVDANRNWKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAVANFLRKH 313

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              |:||:| |:|:| :::|||| :   :   :   | :|| |::      || :|  |
Db        314 RKHIRAYLSFHAYAQMLLYPYSYKYATIPNFRCVESAAYKAVNALQSV-YGVRYRYGPAS 372

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
               |||:: |   || |  || |:|
Db        373 TTLYVSSGSSMDWAYKNGIPYAF 395


RESULT 14
US-09-910-059-113
; Sequence 113, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody,
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 113
;   LENGTH: 613
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised Fd mutant HCPB sequence
US-09-910-059-113

  Query Match             41.0%;  Score 549;  DB 10;  Length 613;
  Best Local Similarity   40.1%;  Pred. No. 1.2e-48;
  Matches  108;  Conservative   47;  Mismatches   74;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| :::: ||::||   :|:||| ||    | ||
Db        307 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 365

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              ::||| |||||||||| ||:                              |:
Db        366 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 425

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        426 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 484

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::|||| :::|||      ::: ||: :|   |: :  :   |:||
Db        485 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 543

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| : |   || || ||:||||
Db        544 YGPGATTIYPSAGTSKDWAYDQGIRYSFT 572


RESULT 15
US-09-910-059-125
; Sequence 125, Application US/09910059
; Patent No. US20020142359A1
; GENERAL INFORMATION:
;  APPLICANT: Copley, Clive G
;  APPLICANT:  Edge, Michael Derek
;  APPLICANT:  Emery, Stephen Charles
;  TITLE OF INVENTION: Monoclonal Antibody to CEA, Conjugates Comprising Said Antibody, a
;  TITLE OF INVENTION:  Their Therapeutic use in an Adept System
;  FILE REFERENCE: 1991-209
;  CURRENT APPLICATION NUMBER: US/09/910,059
;  CURRENT FILING DATE:  2001-07-23
;  PRIOR APPLICATION NUMBER: US 09/171,945
;  PRIOR FILING DATE: 1998-10-29
;  PRIOR APPLICATION NUMBER: PCT/GB97/01165
;  PRIOR FILING DATE: 1997-04-29
;  PRIOR APPLICATION NUMBER: GB 9703103.3
;  PRIOR FILING DATE: 1997-02-14
;  PRIOR APPLICATION NUMBER: GB9609405.7
;  PRIOR FILING DATE: 1996-05-04
;  NUMBER OF SEQ ID NOS: 131
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 125
;   LENGTH: 716
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: humanised pre-pro HCPB-linker-Fd sequence
US-09-910-059-125

  Query Match             41.0%;  Score 549;  DB 10;  Length 716;
  Best Local Similarity   40.1%;  Pred. No. 1.5e-48;
  Matches  108;  Conservative   47;  Mismatches   74;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| :::: ||::||   :|:||| ||    | ||
Db        111 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 169

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              ::||| |||||||||| ||:                              |:
```

```
Db        170 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 229

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  : ||||| |||     ||| ||| : | |||||   ||| | ||:|
Db        230 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 288

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::||||| :::|||     ::: ||: :|   |: :  :   |:||
Db        289 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 347

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| : |   || || ||:||||
Db        348 YGPGATTIYPSAGTSKDWAYDQGIRYSFT 376


Search completed: January  1, 2004, 19:50:59
Job time : 36.1795 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:25:56 ; Search time 16.9478 Seconds
(without alignments)
1395.903 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1239.5	92.6	423	2	A41204	carboxypeptidase B
2	601	44.9	415	2	A32129	carboxypeptidase B
3	586	43.8	306	1	CPBOB	carboxypeptidase B
4	561	41.9	417	1	A34487	carboxypeptidase A
5	552	41.3	309	2	A38395	mast cell carboxyp
6	552	41.3	417	2	A43929	carboxypeptidase A
7	547.5	40.9	416	1	A42332	carboxypeptidase B
8	502.5	37.6	419	1	CPBOA	carboxypeptidase A
9	487.5	36.4	419	1	CPRTA	carboxypeptidase A
10	483	36.1	417	2	A32128	carboxypeptidase A
11	465.5	34.8	419	1	S29127	carboxypeptidase A
12	456	34.1	417	2	A56171	carboxypeptidase A
13	403	30.1	323	2	T20507	hypothetical prote
14	396.5	29.6	528	2	T33527	hypothetical prote
15	375.5	28.1	303	1	CPCYB	carboxypeptidase B
16	366.5	27.4	455	2	T33572	hypothetical prote
17	363	27.1	527	2	T27572	hypothetical prote
18	344.5	25.7	667	2	T33526	hypothetical prote
19	342	25.6	666	2	T24170	hypothetical prote
20	286.5	21.4	424	2	S17571	carboxypeptidase T
21	286.5	21.4	430	2	S48976	ECM14 protein - ye
22	282.5	21.1	451	2	S20723	carboxypeptidase (
23	250.5	18.7	373	2	T26030	hypothetical prote
24	246	18.4	497	2	T40260	carboxypeptidase p
25	177.5	13.3	999	2	T36021	probable zinc-bind
26	139.5	10.4	558	2	S75104	hypothetical prote
27	131.5	9.8	1446	2	T30916	carboxypeptidase D
28	109.5	8.2	985	2	T29910	hypothetical prote
29	101	7.5	1216	2	H84629	hypothetical prote
30	99	7.4	444	2	S51199	tryptophan 5-monoo
31	99	7.4	491	2	B96739	hypothetical prote
32	97	7.2	444	1	S10489	tryptophan 5-monoo
33	96.5	7.2	376	1	E69957	gamma-D-glutamyl-L
34	96	7.2	1389	2	I50090	carboxypeptidase g
35	95	7.1	707	2	T40070	origin recognition
36	94	7.0	477	2	S16383	carboxypeptidase E
37	93.5	7.0	355	2	C83850	gamma-D-glutamyl-L
38	92	6.9	985	2	D82776	pyruvate dehydroge
39	91	6.8	262	2	A64882	probable carboxype
40	91	6.8	262	2	A99867	probable carboxype
41	91	6.8	262	2	H85751	probable carboxype
42	90.5	6.8	448	2	A24327	carboxypeptidase E
43	90.5	6.8	476	2	S12461	carboxypeptidase E
44	90.5	6.8	476	2	A40469	carboxypeptidase E
45	89.5	6.7	1119	2	T13284	carboxypeptidase (

ALIGNMENTS

RESULT 1
A41204
carboxypeptidase B (EC 3.4.17.2) CPB2 precursor - human
N;Alternate names: plasma carboxypeptidase B
C;Species: Homo sapiens (man)
C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 26-Aug-1999
C;Accession: A41204
R;Eaton, D.L.; Malloy, B.E.; Tsai, S.P.; Henzel, W.; Drayna, D.
J. Biol. Chem. 266, 21833-21838, 1991
A;Title: Isolation, molecular cloning, and partial characterization of a novel carboxypep
A;Reference number: A41204; MUID:92042093; PMID:1939207
A;Accession: A41204
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-423 <EAT>
A;Cross-references: GB:M75106; NID:g189686; PID:g189687
C;Genetics:
A;Gene: GDB:CPB2
A;Cross-references: GDB:129546; OMIM:212070
A;Map position: 13q14.11-13q14.11
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zymogen
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-114/Domain: activation peptide #status predicted <ACP>
F;115-423/Product: carboxypeptidase B CPB2 #status predicted <MAT>

Query Match 92.6%; Score 1239.5; DB 2; Length 423;
Best Local Similarity 86.2%; Pred. No. 7.5e-104;
Matches 232; Conservative 0; Mismatches 0; Indels 37; Gaps 1;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383
```

RESULT 2

A32129
carboxypeptidase B (EC 3.4.17.2) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 22-Jun-1999
C;Accession: A32129; S17543
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: A32129
A;Molecule type: DNA
A;Residues: 1-415 <CLA>
A;Cross-references: GB:M23959; GB:J04041; NID:g203293; PIDN:AAA40872.1; PID:g203295
R;Kodama, H.; Shimojo, N.; Suzuki, K.T.
Biochem. J. 278, 857-862, 1991
A;Title: Distribution of manganese in rat pancreas and identification of its primary bin
A;Reference number: S17543; MUID:91378950; PMID:1898371
A;Accession: S17543
A;Molecule type: protein
A;Residues: 109-116,'X',118-130 <KOD>
C;Genetics:
A;Gene: CPB
A;Introns: 22/2; 47/3; 89/2; 122/3; 156/3; 190/3; 227/3; 258/1; 325/3; 354/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;354,376/Active site: Tyr, Glu #status predicted

Query Match 44.9%; Score 601; DB 2; Length 415;
Best Local Similarity 44.6%; Pred. No. 2.5e-46;
Matches 120; Conservative 42; Mismatches 67; Indels 40; Gaps 5;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||   | :|:    | :||: :   :||::|:  ||::||   :||||: || :  | ||
Db        109 ASGHSYTKYNKWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAI 167

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              :|||| ||||||||||| ||:                              |:
Db        168 FIDCGFHAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTK 227

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :||||| ||  | : |:|   |||| :  ||| ||| | |||||||  |||| | ||:|
Db        228 DRMWRKTRSTMAGSSCLGVRPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALAD 286

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:| |:: ||||:::||||| :::||||     :::|||: :   | : :   |   |:||
Db        287 FIRNNLSTIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYT 345

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || |||||||
Db        346 YGPGATTIYPAAGGSDDWSYDQGIKYSFT 374
```

RESULT 3
CPBOB
carboxypeptidase B (EC 3.4.17.2) - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 02-Aug-1994
C;Accession: A93797; A92150; A00912
R;Titani, K.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 72, 1666-1670, 1975
A;Title: Amino-acid sequence of bovine carboxypeptidase B.
A;Reference number: A93797; MUID:75217824; PMID:1057162
A;Accession: A93797
A;Molecule type: protein
A;Residues: 1-306 <TIT>
R;Schmidt, J.J.; Hirs, C.H.W.
J. Biol. Chem. 249, 3756-3764, 1974
A;Title: Primary structure of bovine carboxypeptidase B. Inferences from the locations o
A;Reference number: A92150; MUID:74260705; PMID:4833744
A;Accession: A92150
A;Molecule type: protein
A;Residues: 31-93;131-181;263-265;292-306 <SCH>
R;Schmid, M.F.; Herriott, J.R.
J. Mol. Biol. 103, 175-190, 1976
A;Title: Structure of carboxypeptidase B at 2.8 angstrom resolution.
A;Reference number: A92839; MUID:76265065; PMID:957425
A;Contents: annotation; X-ray crystallography, 2.8 angstroms, and disulfide bonds
R;Plummer Jr., T.H.
J. Biol. Chem. 244, 5246-5253, 1969
A;Title: Isolation and sequence of peptides at the active center of bovine carboxypeptida
A;Reference number: A92051; MUID:70007159; PMID:5344132
A;Contents: annotation; active site
R;Kimmel, M.T.; Plummer Jr., T.H.
J. Biol. Chem. 247, 7864-7869, 1972
A;Title: Identification of a glutamic acid at the active center of bovine carboxypeptidas
A;Reference number: A92119; MUID:73061487; PMID:4565668
A;Contents: annotation; active site
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;63-76,135-158,149-163/Disulfide bonds: #status experimental
F;66,69,194/Binding site: zinc (His, Glu, His) #status experimental
F;246,268/Active site: Tyr, Glu #status experimental

Query Match 43.8%; Score 586; DB 1; Length 306;
Best Local Similarity 43.9%; Pred. No. 3.8e-45;
Matches 116; Conservative 44; Mismatches 64; Indels 40; Gaps 5;

```
Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| | :   :||:::: ||::|   :|:||| ||  : | |:::|||
Db          6 YEKYNNWETIEAWTEQVASENPDLISRSAIGTTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              |||||||||| ||:                              |:      |||||
Db         65 FHAREWISPAFCQWFVREAVRTYGREIHMTEFLDKLDFYVLPVVNIDGYIYTWTTNRMWR 124

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  | : | ||||||||    ||  |||:: |||||||   ||| | |||| |:| :
Db        125 KTRSTRAGSSCTGTDLNRNF-DAGWCSIGASNNPCSETYCGSAAESEKESKAVADFIRNH 183

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::||||| :::||||     |:: ||: :|  ||: :  :   | |::| |:
Db        184 LSSIKAYLTIHSYSQMMLYPYSYDYKLPKNNVELNTLAKGAVKKL-ASLHGTTYSYGPGA 242

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | || ||| || |||||||
Db        243 TTIYPASGGSDDWAYDQGIKYSFT 266
```

RESULT 4
A34487
carboxypeptidase A (EC 3.4.17.1) precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: A34487
R;Reynolds, D.S.; Stevens, R.L.; Gurley, D.S.; Lane, W.S.; Austen, K.F.; Serafin, W.E.
J. Biol. Chem. 264, 20094-20099, 1989
A;Title: Isolation and molecular cloning of mast cell carboxypeptidase A. A novel member
A;Reference number: A34487; MUID:90062123; PMID:2584208
A;Accession: A34487
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:J05118; NID:g192372; PIDN:AAA37369.1; PID:g309135
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

Query Match 41.9%; Score 561; DB 1; Length 417;
Best Local Similarity 43.7%; Pred. No. 1e-42;
Matches 115; Conservative 37; Mismatches 71; Indels 40; Gaps 5;

```
Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|:  ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::|||
Db        116 YAKYNDWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI-------GHN------------------------------RMWR 88
```

```
                ||||||||||||| ||:       | |                                  ||||
Db        175 IHAREWISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
                ||||   |: ||||||||||    |     ::  |   |.|  |||| | ||| :|:| :
Db        235 KNRSRNQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
                :| |||||: ||||| :: || ||     :|::|  ||  |  |:  |   ||| :|  :
Db        294 LNSIKAYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIA 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
                 |:|   |   ||:||||||::|
Db        353 STIYKTSGSSLDWVYDLGIKHTF 375


RESULT 5
A38395
mast cell carboxypeptidase (EC 3.4.-.-) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #text_change 06-Dec-1996
C;Accession: A38395; A33118
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
Biochemistry 30, 648-655, 1991
A;Title: Rat mast cell carboxypeptidase: amino acid sequence and evidence of enzyme acti
A;Reference number: A38395; MUID:91105153; PMID:1988052
A;Accession: A38395
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-309 <COL>
R;Cole, K.R.; Kumar, S.; Le Trong, H.; Woodbury, R.G.; Walsh, K.A.; Neurath, H.
submitted to the Protein Sequence Database, August 1990
A;Reference number: A33118
A;Accession: A33118
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-309 <CO2>
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase
F;248,270/Active site: Tyr, Glu #status predicted

  Query Match             41.3%;  Score 552;  DB 2;  Length 309;
  Best Local Similarity   42.6%;  Pred. No. 4.5e-42;
  Matches  112;  Conservative   39;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
                | :|:  |:| || | : |:||:|:::| |||: |  ||||||: |::   : ||::|||
Db          8 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 66

Qy         66 IHAREWISPAFCLWFI-------GHN-----------------------------RMWR 88
                ||||||:||||| ||:       | |                           ||||
Db         67 IHAREWVSPAFCQWFVYQAAKSYGKNNIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 126

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
                ||||   :: ||||||||||    |     : :  |   | |  |||| | ||| :|:| :
Db        127 KNRSKNPSSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 185

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
                :| |||||: ||||| ::||| ||     :|::|  ||  |   :  :   ||| :|  :
Db        186 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
                 |:|   |   || ||||||::|
Db        245 STIYKTSGSSLDWAYDLGIKHTF 267


RESULT 6
A43929
carboxypeptidase A (EC 3.4.17.1) CPA3 precursor - human
N;Alternate names: mast cell carboxypeptidase A3
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
C;Accession: A43929; A39246; A45759
R;Reynolds, D.S.; Gurley, D.S.; Austen, K.F.
J. Clin. Invest. 89, 273-282, 1992
A;Title: Cloning and characterization of the novel gene for mast cell carboxypeptidase A
A;Reference number: A43929; MUID:92105393; PMID:1729276
A;Accession: A43929
A;Molecule type: DNA
A;Residues: 1-417 <REY>
A;Cross-references: GB:M73716
A;Experimental source: mast cell
A;Note: the authors translated the codon CGC for residue 231 as Thr
R;Reynolds, D.S.; Gurley, D.S.; Stevens, R.L.; Sugarbaker, D.J.; Austen, K.F.; Serafin, W
Proc. Natl. Acad. Sci. U.S.A. 86, 9480-9484, 1989
A;Title: Cloning of cDNAs that encode human mast cell carboxypeptidase A, and comparison
A;Reference number: A39246; MUID:90083291; PMID:2594780
A;Accession: A39246
A;Molecule type: mRNA
A;Residues: 1-417 <RE2>
A;Cross-references: GB:M27717; NID:g179933; PIDN:AAA35652.1; PID:g179934
R;Goldstein, S.M.; Kaempfer, C.E.; Kealey, J.T.; Wintroub, B.U.
J. Clin. Invest. 83, 1630-1636, 1989
A;Title: Human mast cell carboxypeptidase. Purification and characterization.
A;Reference number: A45759; MUID:89214692; PMID:2708524
A;Accession: A45759
A;Molecule type: protein
A;Residues: 110-137 <GOL>
C;Genetics:
A;Gene: GDB:CPA3
A;Cross-references: GDB:125231; OMIM:114851
A;Map position: 3q21.3-3q25
A;Introns: 23/2; 48/3; 90/2; 124/3; 158/3; 192/3; 229/3; 260/1; 327/3; 356/1
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc,
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-109/Domain: activation peptide #status predicted <ACT>
F;110-417/Product: carboxypeptidase A, mast cell #status predicted <MAT>
F;176,179,304/Binding site: zinc (His, Glu, His) #status predicted
F;245-268/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             41.3%;  Score 552;  DB 2;  Length 417;
  Best Local Similarity   41.4%;  Pred. No. 6.5e-42;
  Matches  109;  Conservative   45;  Mismatches   69;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
                | :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||
Db        116 YAKYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI-------------------------------GH------NRMWR 88
                ||||||:||||| ||:                               |:      |||||
Db        175 IHAREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
                ||||   |: ||||||||||  :  |     ::  |:: | |  |||| | ||| :|:| :
Db        235 KNRSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
                :|:|| ||: ||||| ::||| ||      :||:|: ||       :  |   ||| :|
Db        294 LNEIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIE 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
                 |:|   |   || ||||||::|
Db        353 STIYPISGSSLDWAYDLGIKHTF 375


RESULT 7
A42332
carboxypeptidase B (EC 3.4.17.2) precursor, pancreatic - human
N;Alternate names: carboxypeptidase B isozyme 1
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: A42332; S08254; S02813; S02812
R;Yamamoto, K.K.; Pousette, A.; Chow, P.; Wilson, H.; el Shami, S.; French, C.K.
J. Biol. Chem. 267, 2575-2581, 1992
```

```
A;Title: Isolation of a cDNA encoding a human serum marker for acute pancreatitis. Ident
A;Reference number: A42332; MUID:92129345; PMID:1370825
A;Accession: A42332
A;Molecule type: mRNA
A;Residues: 1-416 <YAM>
A;Cross-references: GB:M81057; NID:g809194; PIDN:AAA66973.1; PID:g189625
A;Experimental source: pancreas
A;Note: sequence extracted from NCBI backbone (NCBIN:78573, NCBIP:78574)
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08254
A;Molecule type: protein
A;Residues: 16-40 <MOU>
A;Experimental source: pancreas
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02813
A;Molecule type: protein
A;Residues: 16-36,'X',38-43 <PAS>
A;Experimental source: pancreas
A;Accession: S02812
A;Molecule type: protein
A;Residues: 16-36,'Q',38-43 <PA2>
A;Experimental source: pancreas
C;Genetics:
A;Gene: GDB:CPB1
A;Cross-references: GDB:132160; OMIM:114852
A;Map position: 3q21.3-3q25
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-15/Domain: signal sequence #status predicted <SIG>
F;16-110/Domain: activation peptide #status predicted <ACP>
F;111-415/Product: carboxypeptidase B, pancreatic #status predicted <MAT>
F;173-186,258-272/Disulfide bonds: #status predicted
F;176,179,303/Binding site: zinc (His, Glu, His) #status predicted
F;355,377/Active site: Tyr, Glu #status predicted

  Query Match            40.9%;  Score 547.5;  DB 1;  Length 416;
  Best Local Similarity  40.9%;  Pred. No. 1.6e-41;
  Matches  110;  Conservative   46;  Mismatches   72;  Indels   41;  Gaps    6;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db        111 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 169

Qy         61 WIDCGIHAREWISPAFCLWFI-----------------------------GH------ 83
              ::||| |||||||||| ||:                             |:
Db        170 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 229

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  :  |||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        230 SRFWRKTRSTHTGS-SIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 287

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::|||| :::|||     ::: ||: :|   |: :  :   |:||
Db        288 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 346

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || ||:||||
Db        347 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 375


RESULT 8
CPBOA
carboxypeptidase A (EC 3.4.17.1) precursor [validated] - bovine
N;Alternate names: procarboxypeptidase A activation peptide
C;Species: Bos primigenius taurus (cattle)
C;Date: 24-Apr-1984 #sequence_revision 30-Jun-1992 #text_change 05-May-2000
C;Accession: JN0126; JT0440; A31406; S00059; A38834; A00910; A90355
R;Le Hueerou, I.; Guilloteau, P.; Toullec, R.; Puigserver, A.; Wicker, C.
Biochem. Biophys. Res. Commun. 175, 110-116, 1991
A;Title: Cloning and nucleotide sequence of a bovine pancreatic preprocarboxypeptidase A
A;Reference number: JN0126; MUID:91151335; PMID:1998496
A;Accession: JN0126
A;Molecule type: mRNA
A;Residues: 1-419 <LEH>
A;Cross-references: GB:M61851; NID:g162788; PIDN:AAA30426.1; PID:g162789
R;Foglizzo, E.; Bonicel, J.; Kerfelec, B.; Granon, S.; Chapus, C.
Biochim. Biophys. Acta 954, 183-188, 1988
A;Title: Primary structure of the activation peptide from bovine pancreatic procarboxypep
A;Reference number: JT0440; MUID:88209583; PMID:3365436
A;Accession: JT0440
A;Molecule type: protein
A;Residues: 17-110 <FOG>
R;Wade, R.D.; Hass, G.M.; Kumar, S.; Walsh, K.A.; Neurath, H.
Biochimie 70, 1137-1142, 1988
A;Title: The amino acid sequence of the activation peptide of bovine pro-carboxypeptidase
A;Reference number: A31406; MUID:89150306; PMID:3147705
A;Accession: A31406
A;Molecule type: protein
A;Residues: 17-110 <WAD>
R;Chapus, C.; Kerfelec, B.; Foglizzo, E.; Bonicel, J.
Eur. J. Biochem. 166, 379-385, 1987
A;Title: Further studies on the activation of bovine pancreatic procarboxypeptidase A by
A;Reference number: S00059; MUID:87275921; PMID:3609014
A;Accession: S00059
A;Molecule type: protein
A;Residues: 17-69 <CHA>
R;Bradshaw, R.A.; Ericsson, L.H.; Walsh, K.A.; Neurath, H.
Proc. Natl. Acad. Sci. U.S.A. 63, 1389-1394, 1969
A;Title: The amino acid sequence of bovine carboxypeptidase A.
A;Reference number: A38834; MUID:70050818; PMID:5260942
A;Accession: A38834
A;Molecule type: protein
A;Residues: 111-137,'Q',139-140,'E',142-417 <BRA>
R;Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 961-972, 1971
A;Title: Amino acid sequence of bovine carboxypeptidase A. Isolation and characterizatio
A;Reference number: A90360; MUID:71155185; PMID:4927805
A;Accession: A00910
A;Molecule type: protein
A;Residues: 214-411 <BR2>
A;Note: this is the final paper in a series
R;Petra, P.H.; Hermodson, M.A.; Walsh, K.A.; Neurath, H.
Biochemistry 10, 4023-4025, 1971
A;Title: Characterization of bovine carboxypeptidase A(Allan).
A;Reference number: A90355; MUID:72138789; PMID:5143102
A;Accession: A90355
A;Molecule type: protein
A;Residues: 111-142 <PET>
R;Petra, P.H.; Bradshaw, R.A.; Walsh, K.A.; Neurath, H.
Biochemistry 8, 2762-2768, 1969
A;Title: Identification of the amino acid replacements characterizing the allotropic form
A;Reference number: A90556; MUID:69283620; PMID:5817619
A;Contents: annotation; allelic variant
A;Note: an allelic variant having 179-Val, 228-Glu, and 305-Leu occurs with nearly equal
R;Lipscomb, W.N.; Rees, D.C.
submitted to the Brookhaven Protein Data Bank, March 1982
A;Reference number: A50963; PDB:4CPA
A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 111-137,'Q',139-14
R;Rees, D.C.; Lewis, M.; Lipscomb, W.N.
J. Mol. Biol. 168, 367-387, 1983
A;Title: Refined crystal structure of carboxypeptidase A at 1.54 angstrom resolution.
A;Reference number: A92902; MUID:83294519; PMID:6887246
A;Contents: annotation; X-ray crystallography, 1.54 angstroms
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status experimental <ACP>
F;111-419/Product: carboxypeptidase A #status experimental <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status experimental
F;248-271/Disulfide bonds: #status experimental
```

```
F;358,380/Active site: Tyr, Glu #status experimental

  Query Match            37.6%;  Score 502.5;  DB 1;  Length 419;
  Best Local Similarity  40.2%;  Pred. No. 1.9e-37;
  Matches  106;  Conservative   42;  Mismatches   75;  Indels   41;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:|:||| ::: :   || :::|: || |:|  |:|||| |    : : ||||| |
Db        119 YATYHTLDEIYDFMDLLVAEHPQLVSKLQIGRSYEGRPIYVLKFS-TGGSNRPAIWIDLG 177

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||:||||: |  :||                                    ||:||
Db        178 IHSREWITQATGVWFAKKPTEDYGQDPSFTAILDSMDIFLEIVTNPDGFAFTHSQNRLWR 237

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  ::: |:| | |||:    : : ||||| ||||| | |  || |||::  |:: :
Db        238 KTRSVTSSSLCVGVDANRNW-DAGFGKAGASSSPCSETYHGKYANSEVEVKSIVDFVKDH 296

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              |  ||::|:||||| :::|| ||     |  ||: ||  || |: |:   | | :|
Db        297 GN-FKAFLSIHSYSQLLLYPYGYTTQSIPDKTELNQVAKSAVEAL-KSLYGTSYKYGSII 354

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | ||  || |: |||||||
Db        355 TTIYQASGGSIDWSYNQGIKYSFT 378


RESULT 9
CPRTA
carboxypeptidase A (EC 3.4.17.1) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 13-Jun-1983 #sequence_revision 13-Jun-1983 #text_change 24-Sep-1999
C;Accession: A00911; B32129
R;Quinto, C.; Quiroga, M.; Swain, W.F.; Nikovits Jr., W.C.; Standring, D.N.; Pictet, R.L
Proc. Natl. Acad. Sci. U.S.A. 79, 31-35, 1982
A;Title: Rat preprocarboxypeptidase A: cDNA sequence and preliminary characterization of
A;Reference number: A00911; MUID:82105986; PMID:6275388
A;Accession: A00911
A;Molecule type: mRNA
A;Residues: 1-419 <QUI>
A;Cross-references: GB:J00713; NID:g203364; PIDN:AAA40893.1; PID:g203365
R;Clauser, E.; Gardell, S.J.; Craik, C.S.; MacDonald, R.J.; Rutter, W.J.
J. Biol. Chem. 263, 17837-17845, 1988
A;Title: Structural characterization of the rat carboxypeptidase A1 and B genes. Compara
A;Reference number: A92693; MUID:89034324; PMID:3182872
A;Accession: B32129
A;Molecule type: DNA
A;Residues: 1-260,'F',262,'M',264-346,'K',348-419 <CLA>
C;Genetics:
A;Gene: CPA1
A;Introns: 22/2; 49/3; 127/3; 161/3; 196/2; 232/3; 263/2; 329/3; 358/1
A;Note: the authors translated the codon GTC for residue 196 as Ile
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <PRO>
F;111-419/Product: carboxypeptidase A #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

  Query Match            36.4%;  Score 487.5;  DB 1;  Length 419;
  Best Local Similarity  39.2%;  Pred. No. 4.2e-36;
  Matches  105;  Conservative   40;  Mismatches   74;  Indels   49;  Gaps    6;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:|:||| ::: :   || :::|| ||::||  |::||| |    | : ||||| |
Db        119 YATYHTLDEIYEFMDLLVAEHPQLVSKIQIGNTFEGRPIHVLKFS-TGGTNRPAIWIDTG 177

Qy         66 IHAREWISPAFCLWFIGH-----------------------------------NRMWR 88
              ||:|||:: |  :||                                     |||||
Db        178 IHSREWVTQASGVWFAKKVTKDYGQDPTFTAVLDNMDIFLEIVTNPDGFAYTHKTNRMWR 237

Qy         89 KNRSFYANNHCIGTDLNRNFASKHW----CEEGASSSSCSETYCGLYPESEPEVKAVASF 144
              | ||    : |:| | |||      |      : ||||: ||||| | :| || |||::  |
Db        238 KTRSHTQGSLCVGVDPNRN-----WDAGLGKAGASSNPCSETYRGKFPNSEVEVKSIVDF 292

Qy        145 LRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTH 204
              :  : | |||:||:||||| :::|| ||   : |  ||  :|  || |:  :   | : :
Db        293 VTSHGN-IKAFISIHSYSQLLLYPYGYTSEPAPDQAELDQLAKSAVTAL-TSLHGTEFKY 350

Qy        205 GHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |   :|:| | |   || |  |||||||
Db        351 GSIIDTIYQASGSTIDWTYSQGIKYSFT 378


RESULT 10
A32128
carboxypeptidase A2 (EC 3.4.17.15) precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 22-Jun-1999
C;Accession: A32128
R;Gardell, S.J.; Craik, C.S.; Clauser, E.; Goldsmith, E.J.; Stewart, C.B.; Graf, M.; Rutt
J. Biol. Chem. 263, 17828-17836, 1988
A;Title: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evo
A;Reference number: A32128; MUID:89034323; PMID:3182871
A;Accession: A32128
A;Molecule type: mRNA
A;Residues: 1-417 <GAR>
A;Cross-references: GB:M23719; GB:M23721; GB:J04043; NID:g341029; PIDN:AAA40956.1; PID:g5
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match            36.1%;  Score 483;  DB 2;  Length 417;
  Best Local Similarity  38.2%;  Pred. No. 1.1e-35;
  Matches  109;  Conservative   47;  Mismatches   75;  Indels   54;  Gaps    9;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDC 64
              :| ||:| |||  :: :   :| :::|:::|||||  |: ||| | | :   | |||:|
Db        118 FEAYHTLEEIYQEMDNLVAENPGLVSKVNLGSSFENRPMNVLKFSTGGD---KPAIWLDA 174

Qy         65 GIHAREWISPAFCLW-----------------------FI-------GH------NRMW 87
              |||||||:: |  ||                       |:       |:      ||||
Db        175 GIHAREWVTQATALWTANKIASDYGTDPAITSLLNTLDIFLLPVTNPDGYVFSQTTNRMW 234

Qy         88 RKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRR 147
              || ||  : : |:| | |||:   ::   ||||| ||::| |  | || |||::  |::
Db        235 RKTRSKRSGSGCVGVDPNRNW-DANFGGPGASSSPCSDSYHGPKPNSEVEVKSIVDFIKS 293

Qy        148 NINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHG 207
              :  ::||:|::||||| ::||| |  :|  |  ||  || :| :|: |    | |  |
Db        294 H-GKVKAFITLHSYSQLLMFPYGYKCTKPDDFNELDEVAQKAAQAL-KRLHGTSYKVGPI 351

Qy        208 SETLYLAPGGGDDWIYDLGIKYSFTSN----------PPVEKLLP 242
                :| | ||  || |||||||||              | :::||
Db        352 CSVIYQASGGSIDWAYDLGIKYSFAPELRDTAFYGFLLPAKQILP 396


RESULT 11
S29127
carboxypeptidase A (EC 3.4.17.1) CPA1 precursor - human
N;Alternate names: pancreatic carboxypeptidase A1
C;Species: Homo sapiens (man)
C;Date: 25-Feb-1994 #sequence_revision 19-Jan-1996 #text_change 18-Jun-1999
C;Accession: S29127; A34205; S08253; S02810; S71394; S02811
R;Catasus, L.; Villegas, V.; Pascual, R.; Aviles, F.X.; Wicker-Planquart, C.; Puigserver,
Biochem. J. 287, 299-303, 1992
A;Title: cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1.
A;Reference number: S29127; MUID:93038569; PMID:1417781
A;Accession: S29127
A;Molecule type: mRNA
A;Residues: 1-419 <CAT>
A;Cross-references: EMBL:X67318; NID:g35329; PIDN:CAA47732.1; PID:g35330
R;Stewart, E.A.; Craik, C.S.; Hake, L.; Bowcock, A.M.
Am. J. Hum. Genet. 46, 795-800, 1990
```

```
A;Title: Human carboxypeptidase A identifies a BglII RFLP and maps to 7q31-qter.
A;Reference number: A34205; MUID:90196012; PMID:1969228
A;Accession: A34205
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 330-396 <STE>
A;Note: the authors translated the codon CTG for residue 391 as Val
R;Moulard, M.; Michon, T.; Kerfelec, B.; Chapus, C.
FEBS Lett. 261, 179-183, 1990
A;Title: Further studies on the human pancreatic binary complexes involving procarboxype
A;Reference number: S08253; MUID:90169111; PMID:2307232
A;Accession: S08253
A;Molecule type: protein
A;Residues: 17-43;'XXX',114-135 <MOU>
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidas
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02810
A;Molecule type: protein
A;Residues: 17-42 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 a
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71394
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-419 <LAE>
C;Genetics:
A;Gene: GDB:CPA1; CPA
A;Cross-references: GDB:120597; OMIM:114850
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-419/Product: carboxypeptidase A isozyme 1 #status predicted <MAT>
F;179,182,306/Binding site: zinc (His, Glu, His) #status predicted
F;248-271/Disulfide bonds: #status predicted
F;358,380/Active site: Tyr, Glu #status predicted

  Query Match             34.8%;  Score 465.5;  DB 1;  Length 419;
  Best Local Similarity   37.9%;  Pred. No. 4.1e-34;
  Matches  100;  Conservative   44;  Mismatches   79;  Indels   41;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:| ||| ::: :   :| :::|| ||:::|  |:|||| |    : : ||||| |
Db        119 YATYHTLEEIYDFLDLLVAENPHLVSKIQIGNTYEGRPIYVLKFS-TGGSKRPAIWIDTG 177

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||:|||:: |  :||                                     |||||
Db        178 IHSREWVTQASGVWFAKKITQDYGQDAAFTAILDTLDIFLEIVTNPDGFAFTHSTNRMWR 237

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  | : ||| | |||: :      ||||: ||||| | :  || |||::  |:: :
Db        238 KTRSHTAGSLCIGVDPNRNWDAGFGL-SGASSNPCSETYHGKFANSEVEVKSIVDFVKDH 296

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              | |||:||:||||| :::|| |      | :||  ::  || |:  :   |:: :|
Db        297 GN-IKAFISIHSYSQLLMYPYGYKTEPVPDQDELDQLSKAAVTAL-ASLYGTKFNYGSII 354

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              : :| | |   || |  |||||||
Db        355 KAIYQASGSTIDWTYSQGIKYSFT 378


RESULT 12
A56171
carboxypeptidase A2 (EC 3.4.17.15) precursor - human
N;Alternate names: pancreatic carboxypeptidase A2
C;Species: Homo sapiens (man)
C;Date: 28-Apr-1995 #sequence_revision 11-Aug-1995 #text_change 22-Jun-1999
C;Accession: A56171; S02809; S71395
R;Catasus, L.; Vendrell, J.; Aviles, F.X.; Carreira, S.; Puigserver, A.; Billeter, M.
J. Biol. Chem. 270, 6651-6657, 1995
A;Title: The sequence and conformation of human pancreatic procarboxypeptidase A2. cDNA c
A;Reference number: A56171; MUID:95204457; PMID:7896805
A;Accession: A56171
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-417 <CAT>
A;Cross-references: GB:U19977; NID:g790226; PIDN:AAA74425.1; PID:g790227
A;Note: authors translated the codon AGA for residue 339 as Ser, and AGC for residue 340
R;Pascual, R.; Burgos, F.J.; Salva, M.; Soriano, F.; Mendez, E.; Aviles, F.X.
Eur. J. Biochem. 179, 609-616, 1989
A;Title: Purification and properties of five different forms of human procarboxypeptidase
A;Reference number: S02809; MUID:89153096; PMID:2920728
A;Accession: S02809
A;Molecule type: protein
A;Residues: 'S',18-36,'N',38-43 <PAS>
R;Laethem, R.M.; Blumenkopf, T.A.; Cory, M.; Elwell, L.; Moxham, C.P.; Ray, P.H.; Walton,
Arch. Biochem. Biophys. 332, 8-18, 1996
A;Title: Expression and characterization of human pancreatic preprocarboxypeptidase A1 ar
A;Reference number: S71394; MUID:96400327; PMID:8806703
A;Accession: S71395
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-301,'T',303-338,'SR',341-417 <LAE>
C;Genetics:
A;Gene: GDB:CPA2
A;Cross-references: GDB:125230; OMIM:600688
A;Map position: 7q32-7qter
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; zinc
F;1-16/Domain: signal sequence #status predicted <SIG>
F;17-110/Domain: activation peptide #status predicted <ACP>
F;111-417/Product: carboxypeptidase A2 #status predicted <MAT>
F;177,180,304/Binding site: zinc (His, Glu, His) #status predicted
F;246-269/Disulfide bonds: #status predicted
F;356,378/Active site: Tyr, Glu #status predicted

  Query Match             34.1%;  Score 456;  DB 2;  Length 417;
  Best Local Similarity   36.5%;  Pred. No. 2.9e-33;
  Matches  103;  Conservative   45;  Mismatches   80;  Indels   54;  Gaps    7;

Qy          9 YHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIH 67
              ||:| ||   :: :   || :::|::|||||  |: ||| | | :   | |||:| |||
Db        121 YHTLEEISQEMDNLVAEHPGLVSKVNIGSSFENRPMNVLKFSTGGD---KPAIWLDAGIH 177

Qy         68 AREWISPAFCLWFIG-----------------------------------HNRMWRKN 90
              ||||:: |  ||                                     ||||||
Db        178 AREWVTQATALWTANKIVSDYGKDPSITSILDALDIFLLPVTNPDGYVFSQTKNRMWRKT 237

Qy         91 RSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 150
              ||  : : |:| | |||:    :   ||||: ||::| |    || |||::  |:: :
Db        238 RSKVSGSLCVGVDPNRNW-DAGFGGPGASSNPCSDSYHGPSANSEVEVKSIVDFIKSH-G 295

Qy        151 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSET 210
              ::||:| :||||| ::||| |  :|  | :||| || :| ::: ::   |:|  |
Db        296 KVKAFIILHSYSQLLMFPYGYKCTKLDDFDELSEVAQKAAQSL-RSLHGTKYKVGPICSV 354

Qy        211 LYLAPGGGDDWIYDLGIKYSFTSN----------PPVEKLLP 242
              :| | ||  || || ||||||              |  ::||
Db        355 IYQASGGSIDWSYDYGIKYSFAFELRDTGRYGFLLPARQILP 396


RESULT 13
T20507
hypothetical protein F02D8.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C;Accession: T20507; T24289
R;Matthews, L.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19284
```

```
A;Accession: T20507
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WIL>
A;Cross-references: EMBL:Z78411; PIDN:CAB01647.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone F02D8
R;Wild, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19869
A;Accession: T24289
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-323 <WI2>
A;Cross-references: EMBL:Z78413; PIDN:CAB01666.1; GSPDB:GN00023; CESP:F02D8.4
A;Experimental source: clone T01C3
C;Genetics:
A;Gene: CESP:F02D8.4
A;Map position: 5
A;Introns: 33/3; 67/3; 201/3; 228/1; 257/2
C;Superfamily: carboxypeptidase

  Query Match             30.1%;  Score 403;  DB 2;  Length 323;
  Best Local Similarity   33.5%;  Pred. No. 1.3e-28;
  Matches   89;  Conservative   52;  Mismatches   81;  Indels   44;  Gaps    7;

Qy         18 WIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIHAREWISPAFC 77
Db          3 FLNSLAQQYPNDVKLQNIGNSYEGRSITAVRIA-DDGSSKPIVWIDAGIHAREWISYNVA 61

Qy         78 LWFI--------------------------------GHNRMWRKNRSFYANNHCIGTDL 104
Db         62 LYLIYTIVSQPAYRNLLDSVQLVVVPNTNPDGYEYSRTNDRMWRKTRSRFTNSRCAGADA 121

Qy        105 NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQH 164
Db        122 NRNYPF-YWGTQGVSHSQCSEIFCGSRPQSEPEVLALTNAIIRDEERIKGYIALHSYGQE 180

Qy        165 IVFPYSYT-RSKSKDHEELSLVA---SEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDD 220
Db        181 ILYPWGHTQRTYPTDVQDLIQVGRAMASAIRAV----NNTDYTVVNSGDGLYPAAGASDD 236

Qy        221 WIYDLGIKYSFTSN-PPVEKLLPLSL 245
Db        237 WAKSRGIKYSYTIELSPIDDFTGFSL 262


RESULT 14
T33527
hypothetical protein T06A4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2003
C;Accession: T33527
R;Wu, X.; Tin-Wollam, A.; Ozersky, P.; Wilson, R.
submitted to the EMBL Data Library, October 1998
A;Description: The sequence of C. elegans cosmid T06A4.
A;Reference number: Z21364
A;Accession: T33527
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-528 <WUX>
A;Cross-references: EMBL:AF098994; PIDN:AAC67473.1; GSPDB:GN00019; CESP:T06A4.1
A;Experimental source: strain Bristol N2; clone T06A4
C;Genetics:
A;Gene: CESP:T06A4.1
A;Map position: 1
A;Introns: 22/3; 85/1; 120/2; 182/3; 256/3; 294/1; 390/1; 422/2; 477/2
C;Superfamily: carboxypeptidase

  Query Match             29.6%;  Score 396.5;  DB 2;  Length 528;
  Best Local Similarity   35.5%;  Pred. No. 9e-28;
  Matches   98;  Conservative   43;  Mismatches   80;  Indels   55;  Gaps   10;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQT--AKNAIWID 63
Db        140 YDKYNSLEEIQTEMKRLKKEYPTMITLIDIGQSHENRTLLVMKITGKRNPLGSKISMWID 199

Qy         64 CGIHAREWISPAFCLWFIGH----------------------------------------NR 85
Db        200 AGIHAREWIAPATAM-YIAHELLLGYENDATVAKLMDHIDFYILPVMNPDGYEYSREKNR 258

Qy         86 MWRKNRS-------FYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEV 138
Db        259 MWRKNRSPAKCARQTFSTVCCSGVDLNRNF-DWFWASTGSSSDPCHDTYHGSAAFSEPES 317

Qy        139 KAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTR-SKSKD-HEELSLVASEAVRAIEKT 196
Db        318 QAVRDFLEQNTPE--AFISLHSYSQMWLIPYGHRKQSYPQDYHTGLRPLALRATKALYEL 375

Qy        197 SKNTRYTHGHGSETLYLAPGGGDDWIY-DLGIKYSF 231
Db        376 -YGTKYQVGTGADLMYEASGGSHDWAKGQLKVPYAY 410


RESULT 15
CPCYB
carboxypeptidase B (EC 3.4.17.2) - broad-fingered crayfish
C;Species: Astacus astacus, Astacus fluviatilis (broad-fingered crayfish)
C;Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 06-Dec-1996
C;Accession: A05141; A14178
R;Titani, K.; Ericsson, L.H.; Kumar, S.; Jakob, F.; Neurath, H.; Zwilling, R.
Biochemistry 23, 1245-1250, 1984
A;Title: Amino acid sequence of crayfish (Astacus fluviatilis) carboxypeptidase B.
A;Reference number: A90482
A;Accession: A05141
A;Molecule type: protein
A;Residues: 1-303 <TIT>
C;Superfamily: carboxypeptidase
C;Keywords: hydrolase; metallo-carboxypeptidase; metalloprotein; protein digestion; zinc
F;241,264/Active site: Tyr, Glu #status predicted

  Query Match             28.1%;  Score 375.5;  DB 1;  Length 303;
  Best Local Similarity   33.1%;  Pred. No. 3.5e-26;
  Matches   87;  Conservative   45;  Mismatches   92;  Indels   39;  Gaps    6;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
Db          3 WTSYHDYDEINAWLDSLATDYPELASVEDVGLSYEGRTMKLLKL-GKGGADKPIIFIDGG 61

Qy         66 IHAREWISPAFCLWFIGH----------------------------------NRMWRKNR 91
Db         62 IHAREWIAPSTVTYIVNEFVSNSATYDDILSNVNFYVMPTINPDGYAYTFTDDRLWRKTR 121

Qy         92 SFYAN-NHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 150
Db        122 SETGSVLGCKGADPNRNW-SFHWDEVGASDSPCSDIYAGPEPFSEVEMRNVRDQILEYAA 180

Qy        151 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSET 210
Db        181 NIKVYLTFHSYSQLWMYPWGFTSDLPDDWQDLDTLATNAVDALTAV-HGTRYEIGSSTNT 239

Qy        211 LYLAPGGGDDWIY-DLGIKYSFT 232
Db        240 IYAAAGGSDDWAKGEGGVKYAYT 262


Search completed: January  1, 2004, 19:34:33
Job time : 19.1978 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:03:36 ; Search time 9.75783 Seconds
(without alignments)
1185.568 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	601	44.9	415	1	CBPB_RAT	P19223 rattus norv
2	586	43.8	306	1	CBPB_BOVIN	P00732 bos taurus
3	578	43.2	416	1	CBPB_CANFA	P55261 canis famil
4	575	43.0	401	1	CBPB_PIG	P09955 sus scrofa
5	567	42.4	417	1	CBPB_HUMAN	P15086 homo sapien
6	561	41.9	417	1	CBPC_MOUSE	P15089 mus musculu
7	552	41.3	309	1	CBPC_RAT	P21961 rattus norv
8	552	41.3	417	1	CBPC_HUMAN	P15088 homo sapien
9	502.5	37.6	419	1	CBPA_BOVIN	P00730 bos taurus
10	488.5	36.5	419	1	CBP1_RAT	P00731 rattus norv
11	483	36.1	417	1	CBP2_RAT	P19222 rattus norv
12	465.5	34.8	419	1	CBP1_HUMAN	P15085 homo sapien
13	459	34.3	417	1	CPB2_HUMAN	P48052 homo sapien
14	458	34.2	421	1	CBP4_HUMAN	Q9ui42 homo sapien
15	411.5	30.8	304	1	CBPZ_SIMVI	P42788 simulium vi
16	381.5	28.5	433	1	CBPA_ANOGA	O02350 anopheles g
17	375.5	28.1	303	1	CBPB_ASTFL	P04069 astacus flu
18	286.5	21.4	424	1	CBPT_THEVU	P29068 thermoactin
19	286.5	21.4	430	1	YHT2_YEAST	P38836 saccharomyc
20	282.5	21.1	451	1	CBPS_STRGR	P18143 streptomyce
21	255	19.1	434	1	CBPS_STRCP	P39041 streptomyce
22	118.5	8.9	1377	1	CBPD_MOUSE	O89001 mus musculu
23	115.5	8.6	1380	1	CBPD_HUMAN	O75976 homo sapien
24	103.5	7.7	445	1	TPH1_CHICK	P70080 gallus gall
25	103.5	7.7	734	1	CPXM_HUMAN	Q96sm3 homo sapien
26	99	7.4	444	1	TPH1_HUMAN	P17752 homo sapien
27	98	7.3	722	1	CPXM_MOUSE	Q9z100 mus musculu
28	97	7.2	297	1	Y103_HUMAN	Q15006 homo sapien
29	96.5	7.2	376	1	YQGT_BACSU	P54497 bacillus su
30	95	7.1	444	1	TPH1_RABIT	P17290 oryctolagus
31	95	7.1	707	1	ORC1_SCHPO	P54789 schizosacch
32	91	6.8	262	1	YCJI_ECOLI	P51983 escherichia
33	90.5	6.8	434	1	CBPH_BOVIN	P04836 bos taurus
34	90.5	6.8	476	1	CBPH_RAT	P15087 rattus norv
35	89.5	6.7	476	1	CBPH_MOUSE	Q00493 mus musculu
36	89	6.7	481	1	TPH_XENLA	Q92142 xenopus lae
37	88.5	6.6	476	1	CBPH_HUMAN	P16870 homo sapien
38	86	6.4	382	1	HMBP_DROME	P22809 drosophila
39	86	6.4	447	1	TPH1_MOUSE	P17532 mus musculu
40	85.5	6.4	465	1	STHA_ECO57	Q8x727 escherichia
41	85.5	6.4	465	1	STHA_ECOLI	P27306 escherichia
42	85.5	6.4	1406	1	CBPD_DROME	P42787 drosophila
43	85	6.4	1255	1	PER2_HUMAN	O15055 homo sapien
44	83.5	6.2	319	1	UN30_CAEEL	P52906 caenorhabdi
45	83.5	6.2	444	1	TPH1_RAT	P09810 rattus norv

ALIGNMENTS

RESULT 1
CBPB_RAT
ID CBPB_RAT STANDARD; PRT; 415 AA.
AC P19223;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carboxypeptidase B precursor (EC 3.4.17.2).
GN CPB.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89034324; PubMed=3182872;
RA Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT "Structural characterization of the rat carboxypeptidase A1 and B
RT genes. Comparative analysis of the rat carboxypeptidase gene
RT family.";
RL J. Biol. Chem. 263:17837-17845(1988).
CC -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC peptide + L-lysine(or L-arginine).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M23959; AAA40872.1; -.
DR EMBL; M23947; AAA40872.1; JOINED.
DR EMBL; M23950; AAA40872.1; JOINED.
DR EMBL; M23952; AAA40872.1; JOINED.
DR EMBL; M23953; AAA40872.1; JOINED.
DR EMBL; M23954; AAA40872.1; JOINED.
DR PIR; A32129; A32129.
DR HSSP; P09955; 1NSA.
DR MEROPS; M14.003; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR SMART; SM00631; Zn_pept; 1.
DR PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT SIGNAL 1 13 BY SIMILARITY.
FT PROPEP 14 108 ACTIVATION PEPTIDE.
FT CHAIN 109 415 CARBOXYPEPTIDASE B.
FT METAL 174 174 ZINC (BY SIMILARITY).
FT METAL 177 177 ZINC (BY SIMILARITY).
FT METAL 302 302 ZINC (BY SIMILARITY).

```
FT   ACT_SITE    376    376       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    171    184       BY SIMILARITY.
FT   DISULFID    243    266       BY SIMILARITY.
FT   DISULFID    257    271       BY SIMILARITY.
SQ   SEQUENCE   415 AA;  47515 MW;  8EA06CCADE30B6F2 CRC64;

  Query Match             44.9%;  Score 601;  DB 1;  Length 415;
  Best Local Similarity   44.6%;  Pred. No. 2.7e-48;
  Matches  120;  Conservative   42;  Mismatches   67;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||   | :|:    | :||: :   :||::|:  ||::||   :||||: || :  | ||
Db        109 ASGHSYTKYNKWETIEAWIQQVATDNPDLVTQSVIGTTFEGRNMYVLKI-GKTRPNKPAI 167

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              :|||| ||||||||||| ||:                              |:
Db        168 FIDCGFHAREWISPAFCQWFVREAVRTYNQEIHMKQLLDELDFYVLPVVNIDGYVYTWTK 227

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :||||| ||  | : |:|   |||| :  ||| ||| | |||||||  |||| | ||:|
Db        228 DRMWRKTRSTMAGSSCLGVRPNRNF-NAGWCEVGASRSPCSETYCGPAPESEKETKALAD 286

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:| |:: ||||:::|||| :::||||     :::|||: :   | : :  |   |:||
Db        287 FIRNNLSTIKAYLTIHSYSQMMLYPYSYDYKLPENYEELNALVKGAAKEL-ATLHGTKYT 345

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || |||||||
Db        346 YGPGATTIYPAAGGSDDWSYDQGIKYSFT 374


RESULT 2
CBPB_BOVIN
ID   CBPB_BOVIN     STANDARD;      PRT;   306 AA.
AC   P00732;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B (EC 3.4.17.2).
GN   CPB.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=75217824; PubMed=1057162;
RA   Titani K., Ericsson L.H., Walsh K.A., Neurath H.;
RT   "Amino-acid sequence of bovine carboxypeptidase B.";
RL   Proc. Natl. Acad. Sci. U.S.A. 72:1666-1670(1975).
RN   [2]
RP   SEQUENCE OF 31-93; 131-181; 263-265 AND 292-306.
RX   MEDLINE=74260705; PubMed=4833744;
RA   Schmidt J.J., Hirs C.H.W.;
RT   "Primary structure of bovine carboxypeptidase B. Inferences from the
RT   locations of the half-cystines and identification of the active site
RT   arginine.";
RL   J. Biol. Chem. 249:3756-3764(1974).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS), AND DISULFIDE BONDS.
RX   MEDLINE=76265065; PubMed=957425;
RA   Schmid M.F., Herriott J.R.;
RT   "Structure of carboxypeptidase B at 2.8-A resolution.";
RL   J. Mol. Biol. 103:175-190(1976).
RN   [4]
RP   ACTIVE SITE.
RX   MEDLINE=70007159; PubMed=5344132;
RA   Plummer T.H. Jr.;
RT   "Isolation and sequence of peptides at the active center of bovine
RT   carboxypeptidase B.";
RL   J. Biol. Chem. 244:5246-5253(1969).
RN   [5]
RP   ACTIVE SITE.
RX   MEDLINE=73061487; PubMed=4565668;
RA   Kimmel M.T., Plummer T.H. Jr.;
RT   "Identification of a glutamic acid at the active center of bovine
RT   carboxypeptidase B.";
RL   J. Biol. Chem. 247:7864-7869(1972).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A93797; CPBOB.
DR   PDB; 1CPB; 30-SEP-83.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; 3D-structure.
FT   DISULFID     63     76
FT   DISULFID    135    158
FT   DISULFID    149    163
FT   METAL        66     66       ZINC.
FT   METAL        69     69       ZINC.
FT   METAL       194    194       ZINC.
FT   ACT_SITE    246    246
FT   ACT_SITE    268    268       NUCLEOPHILE.
SQ   SEQUENCE   306 AA;  34612 MW;  C329D2655C44A172 CRC64;

  Query Match             43.8%;  Score 586;  DB 1;  Length 306;
  Best Local Similarity   43.9%;  Pred. No. 4.5e-47;
  Matches  116;  Conservative   44;  Mismatches   64;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| | :   :||:::: ||::|    :|:||| ||  : | |:::|||
Db          6 YEKYNNWETIEAWTEQVASENPDLISRSAIGTTFLGNTIYLLKV-GKPGSNKPAVFMDCG 64

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              |||||||||| ||:                              |:      |||||
Db         65 FHAREWISPAFCQWFVREAVRTYGREIHMTEFLDKLDFYVLPVVNIDGYIYTWTTNRMWR 124

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  | : | ||||||||    ||  |||:: |||||||   ||| | |||| |:| :
Db        125 KTRSTRAGSSCTGTDLNRNF-DAGWCSIGASNNPCSETYCGSAAESEKESKAVADFIRNH 183

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::||||| :::|||||      |:: ||: :|  ||: :  :   | |::| |:
Db        184 LSSIKAYLTIHSYSQMMLYPYSYDYKLPKNNVELNTLAKGAVKKL-ASLHGTTYSYGPGA 242

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | || ||| || |||||||
Db        243 TTIYPASGGSDDWAYDQGIKYSFT 266


RESULT 3
CBPB_CANFA
ID   CBPB_CANFA     STANDARD;      PRT;   416 AA.
AC   P55261;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (47 kDa zymogen granule
DE   membrane associated protein) (ZAP47).
GN   CPB1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Fukuoka S.-I.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
```

```
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D78348; BAA11366.1; -.
DR   HSSP; P09955; 1PBA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     15       BY SIMILARITY.
FT   PROPEP       16    109       ACTIVATION PEPTIDE (BY SIMILARITY).
FT   CHAIN       110    416       CARBOXYPEPTIDASE B.
FT   METAL       175    175       ZINC (BY SIMILARITY).
FT   METAL       178    178       ZINC (BY SIMILARITY).
FT   METAL       303    303       ZINC (BY SIMILARITY).
FT   ACT_SITE    377    377       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    244    267       BY SIMILARITY.
FT   DISULFID    258    272       BY SIMILARITY.
SQ   SEQUENCE   416 AA;  47595 MW;  DFE1992CD52F8AB4 CRC64;

  Query Match             43.2%;  Score 578;  DB 1;  Length 416;
  Best Local Similarity   42.4%;  Pred. No. 3.7e-46;
  Matches  114;  Conservative   46;  Mismatches   69;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:     | :| : :|  :||::::  ||::||   :|:||| ||     | ||
Db        110 ATGHSYEKYNRWETIEAWTQQVTSENPDLISRRSIGTTFEGRTIYLLKV-GKAGQNKPAI 168

Qy         61 WIDCGIHAREWISPAFCLWFI-----------------------------GH------ 83
              ::||| |||||||||| ||:                             |:
Db        169 FMDCGFHAREWISPAFWQWFVREXIRTYGQEIHMTELLDKLDFYVLPVGNIDGYVYTWTK 228

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              |||||| ||     :|:|||  |||    ||: ||| : | |||||   ||| | ||:|:
Db        229 NRMWRKTRSTQVGTNCVGTDPTRNF-DAGWCKIGASRNPCDETYCGPAAESEKETKALAN 287

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:| |:: .||||:::|||||  :::||||    :::: ||: :|   |: :  |   |:||
Db        288 FIRSNLSSIKAYLTIHSYSQMMLYPYSYDYKLTENNAELNALAKATVKEL-ATLHGTKYT 346

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || |||||||
Db        347 YGPGATTIYPAAGGSDDWAYDQGIKYSFT 375


RESULT 4
CBPB_PIG
ID   CBPB_PIG       STANDARD;      PRT;   401 AA.
AC   P09955;
DT   01-MAR-1989 (Rel. 10, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2).
GN   CPB.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE OF 1-103.
RX   MEDLINE=91208150; PubMed=2018774;
RA   Burgos F.J., Salva M., Villegas V., Soriano F., Mendez E.,
RA   Aviles F.X.;
RT   "Analysis of the activation process of porcine procarboxypeptidase B
RT   and determination of the sequence of its activation segment.";
RL   Biochemistry 30:4082-4089(1991).
RN   [2]
RP   PRELIMINARY SEQUENCE OF 1-38.
RX   MEDLINE=85279427; PubMed=4026847;
RA   Aviles F.X., Vendrell J., Burgos F.J., Soriano F., Mendez E.;
RT   "Sequential homologies between procarboxypeptidases A and B from
RT   porcine pancreas.";
RL   Biochem. Biophys. Res. Commun. 130:97-103(1985).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX   MEDLINE=91114690; PubMed=1989878;
RA   Coll M., Guasch A., Aviles F.X., Huber R.;
RT   "Three-dimensional structure of porcine procarboxypeptidase B: a
RT   structural basis of its inactivity.";
RL   EMBO J. 10:1-9(1991).
RN   [4]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE, AND SEQUENCE OF 1-81.
RX   MEDLINE=91027767; PubMed=2223783;
RA   Vendrell J., Wider G., Aviles F.X., Wuethrich K.;
RT   "Sequence-specific 1H NMR assignments and determination of the
RT   secondary structure for the activation domain isolated from
RT   pancreatic procarboxypeptidase B.";
RL   Biochemistry 29:7515-7522(1990).
RN   [5]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=91114693; PubMed=1989879;
RA   Vendrell J., Billeter M., Wider G., Aviles F.X., Wuethrich K.;
RT   "The NMR structure of the activation domain isolated from porcine
RT   procarboxypeptidase B.";
RL   EMBO J. 10:11-15(1991).
RN   [6]
RP   STRUCTURE BY NMR OF ACTIVATION PEPTIDE.
RX   MEDLINE=93044373; PubMed=1422143;
RA   Billeter M., Vendrell J., Wider G., Aviles F.X., Coll M., Guasch A.,
RA   Huber R., Wuethrich K.;
RT   "Comparison of the NMR solution structure with the X-ray crystal
RT   structure of the activation domain from procarboxypeptidase B.";
RL   J. Biomol. NMR 2:1-10(1992).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COB.html".
DR   PDB; 1PBA; 31-OCT-93.
DR   PDB; 1NSA; 24-DEC-97.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen;
KW   3D-structure.
FT   PROPEP        1     95       ACTIVATION PEPTIDE.
FT   CHAIN        96    401       CARBOXYPEPTIDASE B.
FT   DISULFID    158    171
FT   DISULFID    230    253
FT   DISULFID    244    258
FT   METAL       161    161       ZINC.
FT   METAL       164    164       ZINC.
FT   METAL       289    289       ZINC.
FT   ACT_SITE    341    341
FT   ACT_SITE    363    363       NUCLEOPHILE.
FT   TURN          8      9
```

```
FT   STRAND       11     17
FT   HELIX        20     32
FT   STRAND       36     39
FT   HELIX        43     45
FT   STRAND       50     56
FT   HELIX        58     60
FT   HELIX        61     69
FT   TURN         70     72
FT   STRAND       74     79
FT   HELIX        82     87
FT   TURN         88     88
FT   TURN        101    102
FT   HELIX       107    120
FT   TURN        122    124
FT   STRAND      125    132
FT   TURN        134    135
FT   STRAND      138    144
FT   TURN        149    150
FT   STRAND      153    157
FT   TURN        162    163
FT   HELIX       165    181
FT   TURN        182    184
FT   HELIX       186    194
FT   STRAND      196    200
FT   HELIX       205    213
FT   TURN        214    214
FT   TURN        216    217
FT   STRAND      224    224
FT   TURN        226    227
FT   STRAND      231    231
FT   HELIX       235    237
FT   TURN        243    244
FT   STRAND      249    249
FT   TURN        252    253
FT   TURN        255    256
FT   STRAND      257    257
FT   TURN        262    263
FT   HELIX       266    277
FT   TURN        278    281
FT   STRAND      282    289
FT   STRAND      294    297
FT   TURN        307    308
FT   HELIX       309    327
FT   STRAND      332    335
FT   HELIX       336    339
FT   HELIX       347    353
FT   TURN        354    355
FT   STRAND      359    363
FT   TURN        371    372
FT   HELIX       376    378
FT   HELIX       379    398
FT   TURN        399    401
SQ   SEQUENCE   401 AA;  45713 MW;  53129AF159A26348 CRC64;

  Query Match           43.0%;  Score 575;  DB 1;  Length 401;
  Best Local Similarity   43.6%;  Pred. No. 6.7e-46;
  Matches  115;  Conservative   45;  Mismatches   64;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| | :| ::||:::::  ||::|:   :|:||| ||  : | ||::|||
Db        101 YEKYNNWETIEAWTEQVTSKNPDLISRSAIGTTFDGDNIYLLKV-GKPGSNKPAIFMDCG 159

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
               ||||||| ||| ||:                             |:      |||||
Db        160 FHAREWISQAFCQWFVRDAVRTYGYEAHMTEFLDNLDFYVLPVLNIDGYIYTWTKNRMWR 219

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||   | : | ||| |||| :  ||  ||| : |:|||||   ||| | ||:| |:| |
Db        220 KTRSTNAGSSCTGTDPNRNF-NAGWCTVGASVNPCNETYCGSAAESEKETKALADFIRNN 278

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::||||| |::||||    ::  ||: :|  ||: :  :   | |::| ||
Db        279 LSSIKAYLTIHSYSQMILYPYSYDYKLPENDAELNSLAKGAVKEL-ASLYGTSYSYGPGS 337

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
                  |:| | || ||| |: |||||||
Db        338 TTIYPAAGGSDDWAYNQGIKYSFT 361


RESULT 5
CBPB_HUMAN
ID   CBPB_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15086; O60834; Q96BQ8;
DT   01-APR-1990 (Rel. 14, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase B precursor (EC 3.4.17.2) (Pancreas-specific protein)
DE   (PASP).
GN   CPB1 OR CPB OR PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 16-40.
RC   TISSUE=Pancreas;
RX   MEDLINE=92129345; PubMed=1370825;
RA   Yamamoto K.K., Pousette A., Chow P., Wilson H., el Shami S.,
RA   French C.K.;
RT   "Isolation of a cDNA encoding a human serum marker for acute
RT   pancreatitis. Identification of pancreas-specific protein as
RT   pancreatic procarboxypeptidase B.";
RL   J. Biol. Chem. 267:2575-2581(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=98182241; PubMed=9524066;
RA   Aloy P., Catasus L., Villegas V., Reverter D., Vendrell J.,
RA   Aviles F.X.;
RT   "Comparative analysis of the sequences and three-dimensional models
RT   of human procarboxypeptidases A1, A2 and B.";
RL   Biol. Chem. 379:149-155(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Adrenal gland;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 16-43.
RC   TISSUE=Pancreas;
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-Lysine(or L-arginine) + H(2)O =
CC       peptide + L-lysine(or L-arginine).
```

```
CC   -!- TISSUE SPECIFICITY: Pancreas.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M81057; AAA66973.1; -.
DR   EMBL; AJ224866; CAA12163.1; -.
DR   EMBL; BC015338; AAH15338.1; -.
DR   PDB; 1KWM; 28-AUG-02.
DR   MEROPS; M14.003; -.
DR   Genew; HGNC:2299; CPB1.
DR   MIM; 114852; -.
DR   GO; GO:0004180; F:carboxypeptidase activity; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL       1    15
FT   PROPEP      16   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   417       CARBOXYPEPTIDASE B.
FT   METAL      176   176       ZINC (BY SIMILARITY).
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173   186       BY SIMILARITY.
FT   DISULFID   245   268       BY SIMILARITY.
FT   DISULFID   259   273       BY SIMILARITY.
FT   CONFLICT    16    16       H -> A (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    17    17       H -> Q (IN REF. 1; AA SEQUENCE).
FT   CONFLICT    37    37       H -> Q (IN REF. 4).
FT   CONFLICT   208   208       D -> N (IN REF. 2).
FT   CONFLICT   245   245       MISSING (IN REF. 1).
SQ   SEQUENCE   417 AA;  47367 MW;  EBBB98B27F5D5AF9 CRC64;

  Query Match             42.4%;  Score 567;  DB 1;  Length 417;
  Best Local Similarity   41.3%;  Pred. No. 3.9e-45;
  Matches  111;  Conservative   46;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |:   ||:|:    | :| : :   :| ::::  ||::||   :|:||| ||    | ||
Db        111 ATGHSYEKYNKWETIEAWTQQVATENPALISRSVIGTTFEGRAIYLLKV-GKAGQNKPAI 169

Qy         61 WIDCGIHAREWISPAFCLWFI------------------------------GH------ 83
              ::||| |||||||||| ||:                              |:
Db        170 FMDCGFHAREWISPAFCQWFVREAVRTYGREIQVTELLDKLDFYVLPVLNIDGYIYTWTK 229

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :| ||| || :  : ||||| ||||    ||| ||| : | |||||   ||| | ||:|
Db        230 SRFWRKTRSTHTGSSCIGTDPNRNF-DAGWCEIGASRNPCDETYCGPAAESEKETKALAD 288

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:|  :: ||||:::|||| :::||||      ::: ||: :|   |: :  :   |:||
Db        289 FIRNKLSSIKAYLTIHSYSQMMIYPYSYAYKLGENNAELNALAKATVKEL-ASLHGTKYT 347

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || ||:||||
Db        348 YGPGATTIYPAAGGSDDWAYDQGIRYSFT 376


RESULT 6
CBPC_MOUSE
ID   CBPC_MOUSE     STANDARD;      PRT;   417 AA.
AC   P15089;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90062123; PubMed=2584208;
RA   Reynolds D.S., Stevens R.L., Gurley D.S., Lane W.S., Austen K.F.,
RA   Serafin W.E.;
RT   "Isolation and molecular cloning of mast cell carboxypeptidase A. A
RT   novel member of the carboxypeptidase gene family.";
RL   J. Biol. Chem. 264:20094-20099(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J05118; AAA37369.1; -.
DR   PIR; A34487; A34487.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   MGD; MGI:88479; Cpa3.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    15
FT   PROPEP      16   109       ACTIVATION PEPTIDE.
FT   CHAIN      110   417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL      176   176       ZINC (BY SIMILARITY).
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      304   304       ZINC (BY SIMILARITY).
FT   ACT_SITE   378   378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   173   186       BY SIMILARITY.
FT   DISULFID   245   268       BY SIMILARITY.
SQ   SEQUENCE   417 AA;  48790 MW;  A2B300A068D1BA6D CRC64;

  Query Match             41.9%;  Score 561;  DB 1;  Length 417;
  Best Local Similarity   43.7%;  Pred. No. 1.4e-44;
  Matches  115;  Conservative   37;  Mismatches   71;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|:  ::| || | : |:||:|:::| |||: |  ||||||: ||:   : ||::|||
Db        116 YAKYNDWDKIVSWTEKMLEKHPEMVSRIKIGSTVEDNPLYVLKI-GKKDGERKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI-------GHN-----------------------------RMWR 88
              |||||||||||| ||:       | |                             ||||
Db        175 IHAREWISPAFCQWFVYQATKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTQDRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   |: ||||||||||    |     ::  |   | |  |||| | ||| :|:| :
Db        235 KNRSRNQNSTCIGTDLNRNF-DVSWDSSPNTNKPCLNVYRGPAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
```

```
              :| |||||: ||||| :: || ||     :|::|  ||  |  |:  |   ||| :|  :
Db        294 LNSIKAYITFHSYSQMLLIPYGYTFKLPPNHQDLLKVARIATDAL-STRYETRYIYGPIA 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
                |:|   |   ||:||||||::|
Db        353 STIYKTSGSSLDWVYDLGIKHTF 375


RESULT 7
CBPC_RAT
ID   CBPC_RAT       STANDARD;      PRT;   309 AA.
AC   P21961;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-AUG-1991 (Rel. 19, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mast cell carboxypeptidase (EC 3.4.17.1) (RMC-CP) (Carboxypeptidase
DE   A3).
GN   CPA3.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=91105153; PubMed=1988052;
RA   Cole K.R., Kumar S., le Trong H., Woodbury R.G., Walsh K.A.,
RA   Neurath H.;
RT   "Rat mast cell carboxypeptidase: amino acid sequence and evidence of
RT   enzyme activity within mast cell granules.";
RL   Biochemistry 30:648-655(1991).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
DR   PIR; A38395; A38395.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   METAL        68     68       ZINC (BY SIMILARITY).
FT   METAL        71     71       ZINC (BY SIMILARITY).
FT   METAL       196    196       ZINC (BY SIMILARITY).
FT   ACT_SITE    248    248       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    270    270       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID     65     78       BY SIMILARITY.
FT   DISULFID    137    160       BY SIMILARITY.
FT   VARIANT       1      1       MISSING (IN THE MAJOR FORM).
SQ   SEQUENCE   309 AA;  35786 MW;  20330FABC3EE83EF CRC64;

  Query Match             41.3%;  Score 552;  DB 1;  Length 309;
  Best Local Similarity   42.6%;  Pred. No. 6.6e-44;
  Matches  112;  Conservative   39;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|:  |:| || | : |:||:|:::| |||: |  ||||||: |::   : ||::|||
Db          8 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 66

Qy         66 IHAREWISPAFCLWFI-------GHN----------------------------RMWR 88
              ||||||:|||||  ||:       | |                            ||||
Db         67 IHAREWVSPAFCQWFVYQAAKSYGKNNIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 126

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   :: ||||||||||     |     : : |   | |  |||| | ||| :|:| :
Db        127 KNRSKNPSSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 185

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :| |||||: ||||| ::||| ||     :|::|  ||  |   :  :   ||| :|  :
Db        186 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 244

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
                |:|   |   || ||||||::|
Db        245 STIYKTSGSSLDWAYDLGIKHTF 267


RESULT 8
CBPC_HUMAN
ID   CBPC_HUMAN     STANDARD;      PRT;   417 AA.
AC   P15088; Q96E94;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (EC 3.4.17.1) (MC-CPA)
DE   (Carboxypeptidase A3).
GN   CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=90083291; PubMed=2594780;
RA   Reynolds D.S., Gurley D.S., Stevens R.L., Sugarbaker D.J.,
RA   Austen K.F., Serafin W.E.;
RT   "Cloning of cDNAs that encode human mast cell carboxypeptidase A, and
RT   comparison of the protein with mouse mast cell carboxypeptidase A and
RT   rat pancreatic carboxypeptidases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 86:9480-9484(1989).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Mast cells;
RX   MEDLINE=92105393; PubMed=1729276;
RA   Reynolds D.S., Gurley D.S., Austen K.F.;
RT   "Cloning and characterization of the novel gene for mast cell
RT   carboxypeptidase A.";
RL   J. Clin. Invest. 89:273-282(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [4]
RP   SEQUENCE OF 110-417 FROM N.A.
RX   MEDLINE=92333165; PubMed=1629626;
RA   Natsuaki M., Stewart C.B., Vanderslice P., Schwartz L.B., Natsuaki M.,
RA   Wintroub B.U., Rutter W.J., Goldstein S.M.;
RT   "Human skin mast cell carboxypeptidase: functional characterization,
RT   cDNA cloning, and genealogy.";
RL   J. Invest. Dermatol. 99:138-145(1992).
RN   [5]
RP   SEQUENCE OF 110-137.
RX   MEDLINE=89214692; PubMed=2708524;
```

```
RA   Goldstein S.M., Kaempfer C.E., Kealey J.T., Wintroub B.U.;
RT   "Human mast cell carboxypeptidase. Purification and
RT   characterization.";
RL   J. Clin. Invest. 83:1630-1636(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBCELLULAR LOCATION: Secretory granules.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M27717; AAA35652.1; -.
DR   EMBL; M73720; AAA59568.1; -.
DR   EMBL; M73716; AAA59568.1; JOINED.
DR   EMBL; M73717; AAA59568.1; JOINED.
DR   EMBL; M73718; AAA59568.1; JOINED.
DR   EMBL; M73719; AAA59568.1; JOINED.
DR   EMBL; BC012613; AAH12613.1; -.
DR   EMBL; S40234; AAB22578.2; ALT_SEQ.
DR   PIR; A43929; A43929.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.010; -.
DR   Genew; HGNC:2298; CPA3.
DR   MIM; 114851; -.
DR   GO; GO:0030141; C:secretory granule; NAS.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     15
FT   PROPEP       16    109       ACTIVATION PEPTIDE.
FT   CHAIN       110    417       MAST CELL CARBOXYPEPTIDASE A.
FT   METAL       176    176       ZINC (BY SIMILARITY).
FT   METAL       179    179       ZINC (BY SIMILARITY).
FT   METAL       304    304       ZINC (BY SIMILARITY).
FT   ACT_SITE    378    378       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    173    186       BY SIMILARITY.
FT   DISULFID    245    268       BY SIMILARITY.
FT   CONFLICT     63     63       N -> K (IN REF. 3).
SQ   SEQUENCE   417 AA;  48700 MW;  8CB90DB758117B24 CRC64;

  Query Match             41.3%;  Score 552;  DB 1;  Length 417;
  Best Local Similarity   41.4%;  Pred. No. 9.6e-44;
  Matches  109;  Conservative   45;  Mismatches   69;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|::  :| :| | : :::|:|:::| |||: |  ||||||: |::   : ||::|||
Db        116 YAKYNNWEKIVAWTEKMMDKYPEMVSRIKIGSTVEDNPLYVLKI-GEKNERRKAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              ||||||:|||||  ||:                             |:      |||||
Db        175 IHAREWVSPAFCQWFVYQATKTYGRNKIMTKLLDRMNFYILPVFNVDGYIWSWTKNRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   |: |||||||||| :  |     ::  |:: | |  |||| | ||| :|:| :
Db        235 KNRSKNQNSKCIGTDLNRNF-NASWNSIPNTNDPCADNYRGSAPESEKETKAVTNFIRSH 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :|:|| ||: |||||  ::||| ||     :||:|: ||      :  |   ||| :|
Db        294 LNEIKVYITFHSYSQMLLFPYGYTSKLPPNHEDLAKVAKIGTDVL-STRYETRYIYGPIE 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
                :|:|   |   || ||||||::|
Db        353 STIYPISGSSLDWAYDLGIKHTF 375


RESULT 9
CBPA_BOVIN
ID   CBPA_BOVIN     STANDARD;      PRT;   419 AA.
AC   P00730;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A precursor (EC 3.4.17.1).
GN   CPA.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91151335; PubMed=1998496;
RA   le Hueerou I., Guilloteau P., Toullec R., Puigserver A., Wicker C.;
RT   "Cloning and nucleotide sequence of a bovine pancreatic
RT   preprocarboxypeptidase A cDNA.";
RL   Biochem. Biophys. Res. Commun. 175:110-116(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=96096549; PubMed=8522204;
RA   Goo J.H., Kim K.H., Choi K.Y.;
RT   "Cloning, sequencing and expression of the gene encoding a major
RT   allotypic preprocarboxypeptidase A from bovine pancreas.";
RL   Gene 165:333-334(1995).
RN   [3]
RP   SEQUENCE OF 111-417.
RX   MEDLINE=71155183; PubMed=5102489;
RA   Bradshaw R.A., Walsh K.A., Neurath H.;
RT   "Amino acid sequence of bovine carboxypeptidase A. Tryptic and
RT   chymotryptic peptides of the cyanogen bromide fragment P-I.";
RL   Biochemistry 10:938-950(1971).
RN   [4]
RP   REVISIONS TO 138 AND 141.
RX   MEDLINE=72138789; PubMed=5143102;
RA   Petra P.H., Hermodson M.A., Walsh K.A., Neurath H.;
RT   "Characterization of bovine carboxypeptidase A (Allan).";
RL   Biochemistry 10:4023-4025(1971).
RN   [5]
RP   SEQUENCE OF 17-120.
RX   MEDLINE=89150306; PubMed=3147705;
RA   Wade R.D., Hass G.M., Kumar S., Walsh K.A., Neurath H.;
RT   "The amino acid sequence of the activation peptide of bovine pro-
RT   carboxypeptidase A.";
RL   Biochimie 70:1137-1142(1988).
RN   [6]
RP   X-RAY CRYSTALLOGRAPHY (1.54 ANGSTROMS).
RX   MEDLINE=83294519; PubMed=6887246;
RA   Rees D.C., Lewis M., Lipscomb W.N.;
RT   "Refined crystal structure of carboxypeptidase A at 1.54-A
RT   resolution.";
RL   J. Mol. Biol. 168:367-387(1983).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF TERNARY COMPLEX.
RX   MEDLINE=96003618; PubMed=7556081;
RA   Gomis-Ruth F.X., Gomez M., Bode W., Huber R., Aviles F.X.;
RT   "The three-dimensional structure of the native ternary complex of
RT   bovine pancreatic procarboxypeptidase A with proproteinase E and
RT   chymotrypsinogen C.";
RL   EMBO J. 14:4387-4394(1995).
RN   [8]
RP   VARIANT ALLELIC.
RX   MEDLINE=69283620; PubMed=5817619;
RA   Petra P.H., Bradshaw R.A., Walsh K.A., Neurath H.;
```

```
RT   "Identification of the amino acid replacements characterizing the
RT   allotypic forms of bovine carboxypeptidase A.";
RL   Biochemistry 8:2762-2768(1969).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: MONOMER. THE ZYMOGEN IS SECRETED AS A TERNARY COMPLEX
CC       COMPOSED OF PROCARBOXYPEPTIDASE A, CHYMOTRYPSINOGEN C AND
CC       PROPROTEINASE E.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   -!- DATABASE: NAME=Worthington enzyme manual;
CC       WWW="http://www.worthington-biochem.com/manual/C/COA.html".
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M61851; AAA30426.1; -.
DR   EMBL; M61851; AAA30427.1; -.
DR   EMBL; Z33906; CAA83955.1; -.
DR   PIR; JN0126; CPBOA.
DR   PDB; 3CPA; 15-JAN-87.
DR   PDB; 4CPA; 22-OCT-84.
DR   PDB; 5CPA; 15-JAN-87.
DR   PDB; 6CPA; 15-OCT-91.
DR   PDB; 7CPA; 31-JAN-94.
DR   PDB; 8CPA; 31-JAN-94.
DR   PDB; 1CBX; 31-JAN-94.
DR   PDB; 1CPS; 15-OCT-94.
DR   PDB; 2CTB; 31-JAN-94.
DR   PDB; 2CTC; 31-JAN-94.
DR   PDB; 1ARL; 01-AUG-96.
DR   PDB; 1ARM; 17-AUG-96.
DR   PDB; 1BAV; 01-APR-97.
DR   PDB; 1YME; 12-FEB-97.
DR   PDB; 1CPX; 05-AUG-98.
DR   PDB; 1PYT; 27-JAN-97.
DR   PDB; 1ELL; 26-JUN-02.
DR   PDB; 1ELM; 26-JUN-02.
DR   PDB; 1F57; 06-SEP-00.
DR   PDB; 1HDQ; 15-NOV-01.
DR   PDB; 1HDU; 15-NOV-01.
DR   PDB; 1HEE; 23-NOV-01.
DR   PDB; 1M4L; 04-MAR-03.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure; Polymorphism.
FT   SIGNAL       1     16
FT   PROPEP      17    110       ACTIVATION PEPTIDE.
FT   CHAIN      111    419       CARBOXYPEPTIDASE A.
FT   METAL      179    179       ZINC.
FT   METAL      182    182       ZINC.
FT   METAL      306    306       ZINC.
FT   ACT_SITE   358    358       PROTON DONOR.
FT   ACT_SITE   380    380       NUCLEOPHILE.
FT   DISULFID   248    271
FT   VARIANT    289    289       I -> V (IN ALLELIC VARIANT).
FT   VARIANT    338    338       E -> A (IN ALLELIC VARIANT).
FT   VARIANT    415    415       L -> V (IN ALLELIC VARIANT).
FT   CONFLICT    95     95       S -> L (IN REF. 5).
FT   CONFLICT   199    199       D -> N (IN REF. 3).
FT   CONFLICT   203    203       D -> N (IN REF. 3).
FT   CONFLICT   224    224       D -> N (IN REF. 3).
FT   CONFLICT   232    232       Q -> E (IN REF. 3).
FT   CONFLICT   295    295       D -> N (IN REF. 3).
FT   TURN       114    116
FT   TURN       119    120
FT   HELIX      125    138
FT   TURN       140    142
FT   STRAND     143    150
FT   TURN       152    153
FT   STRAND     156    162
FT   STRAND     171    176
FT   TURN       180    181
FT   HELIX      183    199
FT   TURN       200    202
FT   HELIX      204    212
FT   STRAND     214    218
FT   HELIX      223    231
FT   TURN       232    232
FT   TURN       234    235
FT   STRAND     242    242
FT   TURN       244    245
FT   STRAND     249    249
FT   TURN       253    254
FT   TURN       261    262
FT   STRAND     267    267
FT   TURN       270    271
FT   TURN       273    274
FT   STRAND     275    275
FT   TURN       280    281
FT   HELIX      284    296
FT   STRAND     299    307
FT   STRAND     311    314
FT   TURN       324    325
FT   HELIX      326    344
FT   STRAND     349    352
FT   HELIX      353    356
FT   TURN       357    357
FT   HELIX      364    370
FT   TURN       371    372
FT   STRAND     375    381
FT   TURN       388    389
FT   HELIX      393    395
FT   HELIX      396    416
SQ   SEQUENCE   419 AA;  47082 MW;  21B86407B3BFC452 CRC64;

  Query Match           37.6%;  Score 502.5;  DB 1;  Length 419;
  Best Local Similarity  40.2%;  Pred. No. 3.8e-39;
  Matches  106;  Conservative   42;  Mismatches   75;  Indels   41;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:|:||| ::: :   || :::|: || |:|  |:|||| |    : : ||||| |
Db        119 YATYHTLDEIYDFMDLLVAEHPQLVSKLQIGRSYEGRPIYVLKFS-TGGSNRPAIWIDLG 177

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||:||||: |  :||                                     ||:||
Db        178 IHSREWITQATGVWFAKKFTEDYGQDPSFTAILDSMDIFLEIVTNPDGFAFTHSQNRLWR 237

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  ::: |:| | |||:    : : ||||| ||||| | |  || |||::  |:: :
Db        238 KTRSVTSSSLCVGVDANRNW-DAGFGKAGASSSPCSETYHGKYANSEVEVKSIVDFVKDH 296

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              |  ||::|:||||| :::|| ||      |  ||: ||  || |: |:   | | :|
Db        297 GN-FKAFLSIHSYSQLLLYPYGYTTQSIPDKTELNQVAKSAVEAL-KSLYGTSYKYGSII 354

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
                |:| | ||  || |: |||||||
Db        355 TTIYQASGGSIDWSYNQGIKYSFT 378


RESULT 10
CBP1_RAT
ID   CBP1_RAT       STANDARD;      PRT;   419 AA.
```

```
AC   P00731;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=82105986; PubMed=6275388;
RA   Quinto C., Quiroga M., Swain W.F., Nikovits W.C. Jr., Standring D.N.,
RA   Pictet R.L., Valenzuela P., Rutter W.J.;
RT   "Rat preprocarboxypeptidase A: cDNA sequence and preliminary
RT   characterization of the gene.";
RL   Proc. Natl. Acad. Sci. U.S.A. 79:31-35(1982).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89034324; PubMed=3182872;
RA   Clauser E., Gardell S.J., Craik C.S., Macdonald R.J., Rutter W.J.;
RT   "Structural characterization of the rat carboxypeptidase A1 and B
RT   genes. Comparative analysis of the rat carboxypeptidase gene
RT   family.";
RL   J. Biol. Chem. 263:17837-17845(1988).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; V01232; CAA24542.1; -.
DR   EMBL; J00713; AAA40893.1; -.
DR   EMBL; M23990; AAA40955.1; -.
DR   EMBL; M23960; AAA40955.1; JOINED.
DR   EMBL; M23985; AAA40955.1; JOINED.
DR   EMBL; M23986; AAA40955.1; JOINED.
DR   EMBL; M23987; AAA40955.1; JOINED.
DR   EMBL; M23988; AAA40955.1; JOINED.
DR   EMBL; M23989; AAA40955.1; JOINED.
DR   PIR; A00911; CPRTA.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A1.
FT   METAL      179   179       ZINC (BY SIMILARITY).
FT   METAL      182   182       ZINC (BY SIMILARITY).
FT   METAL      306   306       ZINC (BY SIMILARITY).
FT   ACT_SITE   358   358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE   380   380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID   248   271       BY SIMILARITY.
FT   CONFLICT   196   196       I -> V (IN REF. 1).
FT   CONFLICT   261   263       FGM -> LGK (IN REF. 1).
FT   CONFLICT   347   347       K -> E (IN REF. 1).
SQ   SEQUENCE   419 AA;  47197 MW;  BB002D1CB99B7491 CRC64;
```

```
  Query Match           36.5%;  Score 488.5;  DB 1;  Length 419;
  Best Local Similarity  39.4%;  Pred. No. 7.6e-38;
  Matches  104;  Conservative   42;  Mismatches   77;  Indels   41;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:|:||| ::: :   || :::|| ||::||  |::||| |    | : ||||| |
Db        119 YATYHTLDEIYEFMDLLVAEHPQLVSKIQIGNTFEGRPIHVLKFS-TGGTNRPAIWIDTG 177

Qy         66 IHAREWISPAFCLWFIGH-----------------------------------NRMWR 88
              ||:|||:: |  :||                                       |||||
Db        178 IHSREWVTQASGVWFAKKITKDYGQDPTFTAVLDNMDIFLEIVTNPDGFAYTHKTNRMWR 237

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||    : |:| | |||:    :   ||||: ||||| | :| || |||::  |:  :
Db        238 KTRSHTQGSLCVGVDPNRNW-DAGFGMAGASSNPCSETYRGKFPNSEVEVKSIVDFVTSH 296

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              | |||:||:||||| :::|| ||   : |  ||  :|  || |:  :   |:: :|
Db        297 GN-IKAFISIHSYSQLLLYPYGYTSEPAPDQAELDQLAKSAVTAL-TSLHGTKFKYGSII 354

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              :|:| | |   || |  |||||||
Db        355 DTIYQASGSTIDWTYSQGIKYSFT 378
```

```
RESULT 11
CBP2_RAT
ID   CBP2_RAT       STANDARD;      PRT;   417 AA.
AC   P19222;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=89034323; PubMed=3182871;
RA   Gardell S.J., Craik C.S., Clauser E., Goldsmith E.J., Stewart C.-B.,
RA   Graf M., Rutter W.J.;
RT   "A novel rat carboxypeptidase, CPA2: characterization, molecular
RT   cloning, and evolutionary implications on substrate specificity in
RT   the carboxypeptidase gene family.";
RL   J. Biol. Chem. 263:17828-17836(1988).
RN   [2]
RP   SEQUENCE OF 131-143 FROM N.A.
RX   MEDLINE=95386501; PubMed=7657630;
RA   Normant E., Gros C., Schwartz J.C.;
RT   "Carboxypeptidase A isoforms produced by distinct genes or
RT   alternative splicing in brain and other extrapancreatic tissues.";
RL   J. Biol. Chem. 270:20543-20549(1995).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX   MEDLINE=92105124; PubMed=1761558;
RA   Faming Z., Kobe B., Stewart C.-B., Rutter W.J., Goldsmith E.J.;
RT   "Structural evolution of an enzyme specificity. The structure of rat
RT   carboxypeptidase A2 at 1.9-A resolution.";
RL   J. Biol. Chem. 266:24606-24612(1991).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M23721; AAA40956.1; -.
DR   EMBL; M23714; AAA40956.1; JOINED.
DR   EMBL; M23715; AAA40956.1; JOINED.
DR   EMBL; M23716; AAA40956.1; JOINED.
DR   EMBL; M23717; AAA40956.1; JOINED.
DR   EMBL; M23718; AAA40956.1; JOINED.
DR   EMBL; M23719; AAA40956.1; JOINED.
DR   EMBL; M23720; AAA40956.1; JOINED.
DR   EMBL; S79837; -; NOT_ANNOTATED_CDS.
DR   PIR; A32128; A32128.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.002; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   112       ACTIVATION PEPTIDE.
FT   CHAIN      113   417       CARBOXYPEPTIDASE A2.
FT   METAL      177   177       ZINC.
FT   METAL      180   180       ZINC.
FT   METAL      304   304       ZINC.
FT   ACT_SITE   378   378       NUCLEOPHILE.
FT   DISULFID   246   269       BY SIMILARITY.
FT   DISULFID   318   352       PROBABLE.
SQ   SEQUENCE   417 AA;  46912 MW;  BEDCC41A830F2D45 CRC64;

  Query Match             36.1%;  Score 483;  DB 1;  Length 417;
  Best Local Similarity   38.2%;  Pred. No. 2.5e-37;
  Matches  109;  Conservative   47;  Mismatches   75;  Indels   54;  Gaps    9;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDC 64
              :| ||:| |||  :: :   :| :::|:::||||||  |: ||| | | :   | |||:|
Db        118 FEAYHTLEEIYQEMDNLVAENPGLVSKVNLGSSFENRPMNVLKFSTGGD---KPAIWLDA 174

Qy         65 GIHAREWISPAFCLW------------------------FI--------GH------NRMW 87
              |||||||:: |  ||                        |:        |:      ||||
Db        175 GIHAREWVTQATALWTANKIASDYGTDPAITSLLNTLDIFLLPVTNPDGYVFSQTTNRMW 234

Qy         88 RKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRR 147
              || ||  : : |:| | |||:   ::   ||||| ||::| |  | || |||::  |::
Db        235 RKTRSKRSGSGCVGVDPNRNW-DANFGGPGASSSPCSDSYHGPKPNSEVEVKSIVDFIKS 293

Qy        148 NINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHG 207
              :  ::||:|::||||| ::||| |  :|  |  ||  || :| :|: |    | |  |
Db        294 H-GKVKAFITLHSYSQLLMFPYGYKCTKPDDFNELDEVAQKAAQAL-KRLHGTSYKVGPI 351

Qy        208 SETLYLAPGGGDDWIYDLGIKYSFTSN----------PPVEKLLP 242
                :| | ||  || |||||||||              | :::||
Db        352 CSVIYQASGGSIDWAYDLGIKYSFAFELRDTAFYGFLLPAKQILP 396


RESULT 12
CBP1_HUMAN
ID   CBP1_HUMAN     STANDARD;      PRT;   419 AA.
AC   P15085; Q9BS67;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A1 precursor (EC 3.4.17.1).
GN   CPA1 OR CPA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=93038569; PubMed=1417781;
RA   Catasus L., Villegas V., Pascual R., Aviles F.X.,
RA   Wicker-Planquart C., Puigserver A.;
RT   "cDNA cloning and sequence analysis of human pancreatic
RT   procarboxypeptidase A1.";
RL   Biochem. J. 287:299-303(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 17-42.
RX   MEDLINE=89153096; PubMed=2920728;
RA   Pascual R., Burgos F.J., Soriano F., Mendez E., Aviles F.X.;
RT   "Purification and properties of five different forms of human
RT   procarboxypeptidases.";
RL   Eur. J. Biochem. 179:609-616(1989).
CC   -!- CATALYTIC ACTIVITY: Peptidyl-L-amino acid + H(2)O = peptide + L-
CC       amino acid.
CC   -!- SUBUNIT: Monomer.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67318; CAA47732.1; -.
DR   EMBL; BC005279; AAH05279.1; -.
DR   PIR; S29127; S29127.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.001; -.
DR   Genew; HGNC:2296; CPA1.
DR   MIM; 114850; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL       1    16
FT   PROPEP      17   110       ACTIVATION PEPTIDE.
FT   CHAIN      111   419       CARBOXYPEPTIDASE A1.
```

```
FT   METAL       179    179       ZINC (BY SIMILARITY).
FT   METAL       182    182       ZINC (BY SIMILARITY).
FT   METAL       306    306       ZINC (BY SIMILARITY).
FT   ACT_SITE    358    358       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    380    380       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    248    271       BY SIMILARITY.
FT   CONFLICT    276    276       H -> R (IN REF. 2).
SQ   SEQUENCE   419 AA;  47140 MW;  439FAFFFAEE958B1 CRC64;

  Query Match             34.8%;  Score 465.5;  DB 1;  Length 419;
  Best Local Similarity   37.9%;  Pred. No. 1e-35;
  Matches  100;  Conservative   44;  Mismatches   79;  Indels   41;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:| ||| ::: :   :| :::|| ||:::|  |:|||| |    : : ||||| |
Db        119 YATYHTLEEIYDFLDLLVAENPHLVSKIQIGNTYEGRPIYVLKFS-TGGSKRPAIWIDTG 177

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||:|||:: |  :||                                     |||||
Db        178 IHSREWVTQASGVWFAKKITQDYGQDAAFTAILDTLDIFLEIVTNPDGFAFTHSTNRMWR 237

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||   | : ||| | |||: :      ||||: ||||| | :  || |||::  |:: :
Db        238 KTRSHTAGSLCIGVDPNRNWDAGFGL-SGASSNPCSETYHGKFANSEVEVKSIVDFVKDH 296

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              | |||:||:||||| :::|| |      | :||  ::  || |:  :   |:: :|
Db        297 GN-IKAFISIHSYSQLLMYPYGYKTEPVPDQDELDQLSKAAVTAL-ASLYGTKFNYGSII 354

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              : :| | |   || |  |||||||
Db        355 KAIYQASGSTIDWTYSQGIKYSFT 378


RESULT 13
CPB2_HUMAN
ID   CPB2_HUMAN     STANDARD;      PRT;   417 AA.
AC   P48052; Q96A12; Q96QN3;
DT   01-FEB-1996 (Rel. 33, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A2 precursor (EC 3.4.17.15).
GN   CPA2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND 3D-STRUCTURE MODELING.
RC   TISSUE=Pancreas;
RX   MEDLINE=95204457; PubMed=7896805;
RA   Catasus L., Vendrell J., Aviles F.X., Carreira S., Puigserver A.,
RA   Billeter M.;
RT   "The sequence and conformation of human pancreatic
RT   procarboxypeptidase A2. cDNA cloning, sequence analysis, and
RT   three-dimensional model.";
RL   J. Biol. Chem. 270:6651-6657(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Pancreas;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98111000; PubMed=9450539;
RA   Reverter D., Garcia-Saez I., Catasus L., Vendrell J., Coll M.,
RA   Aviles F.X.;
RT   "Characterisation and preliminary X-ray diffraction analysis of human
RT   pancreatic procarboxypeptidase A2.";
RL   FEBS Lett. 420:7-10(1997).
RN   [4]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=98046021; PubMed=9384570;
RA   Garcia-Saez I., Reverter D., Vendrell J., Aviles F.X., Coll M.;
RT   "The three-dimensional structure of human procarboxypeptidase A2.
RT   Deciphering the basis of the inhibition, activation and intrinsic
RT   activity of the zymogen.";
RL   EMBO J. 16:6906-6913(1997).
CC   -!- CATALYTIC ACTIVITY: Similar to that of carboxypeptidase A
CC       (EC 3.4.17.1), but with a preference for bulkier C-terminal
CC       residues.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U19977; AAA74425.1; -.
DR   EMBL; BC007009; AAH07009.1; -.
DR   EMBL; BC014571; AAH14571.1; -.
DR   EMBL; BC015140; AAH15140.1; -.
DR   PIR; A56171; A56171.
DR   PDB; 1AYE; 13-JAN-99.
DR   PDB; 1DTD; 12-JUL-00.
DR   PDB; 1O6X; 30-JAN-03.
DR   MEROPS; M14.002; -.
DR   Genew; HGNC:2297; CPA2.
DR   MIM; 600688; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; TAS.
DR   GO; GO:0007039; P:vacuolar protein catabolism; TAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal;
KW   3D-structure.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17    112       ACTIVATION PEPTIDE.
FT   CHAIN       113    417       CARBOXYPEPTIDASE A2.
FT   METAL       177    177       ZINC.
FT   METAL       180    180       ZINC.
FT   METAL       304    304       ZINC.
FT   ACT_SITE    378    378       NUCLEOPHILE.
FT   DISULFID    246    269
FT   DISULFID    318    352
FT   CONFLICT     80     80       E -> G (IN REF. 2; AAH07009).
FT   CONFLICT    302    302       T -> I (IN REF. 1).
SQ   SEQUENCE   417 AA;  46828 MW;  3D9C073A98700A83 CRC64;
```

```
  Query Match           34.3%;  Score 459;  DB 1;  Length 417;
  Best Local Similarity   36.5%;  Pred. No. 4.2e-35;
  Matches  103;  Conservative   46;  Mismatches   79;  Indels   54;  Gaps    7;

Qy          9 YHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVS-GKEQTAKNAIWIDCGIH 67
              ||:| ||   :: :   || :::|::|||||  |: ||| | | :   | |||:| |||
Db        121 YHTLEEISQEMDNLVAEHPGLVSKVNIGSSFENRPMNVLKFSTGGD---KPAIWLDAGIH 177

Qy         68 AREWISPAFCLWFIG-----------------------------------HNRMWRKN 90
              ||||:: |  ||                                     ||||||
Db        178 AREWVTQATALWTANKIVSDYGKDPSITSILDALDIFLLPVTNPDGYVFSQTKNRMWRKT 237

Qy         91 RSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 150
              ||  : : |:| | |||:    :   ||||: ||::| |    || |||::  |:: :
Db        238 RSKVSGSLCVGVDPNRNW-DAGFGGPGASSNPCSDSYHGPSANSEVEVKSIVDFIKSH-G 295

Qy        151 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSET 210
              ::||:|::||||| ::||| |  :|  | :||| || :| ::: ::   |:|  |
Db        296 KVKAFITLHSYSQLLMFPYGYKCTKLDDFDELSEVAQKAAQSL-RSLHGTKYKVGPICSV 354

Qy        211 LYLAPGGGDDWIYDLGIKYSFTSN----------PPVEKLLP 242
              :| | ||  || || ||||||             |  ::||
Db        355 IYQASGGSIDWSYDYGIKYSFAFELRDTGRYGFLLPARQILP 396


RESULT 14
CBP4_HUMAN
ID   CBP4_HUMAN     STANDARD;      PRT;   421 AA.
AC   Q9UI42;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Carboxypeptidase A4 precursor (EC 3.4.17.-) (Carboxypeptidase A3).
GN   CPA4 OR CPA3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99310537; PubMed=10383164;
RA   Huang H., Reed C.P., Zhang J.S., Shridhar V., Wang L., Smith D.I.;
RT   "Carboxypeptidase A3 (CPA3): a novel gene highly induced by histone
RT   deacetylase inhibitors during differentiation of prostate epithelial
RT   cancer cells.";
RL   Cancer Res. 59:2981-2988(1999).
RN   [2]
RP   IMPRINTING.
RX   MEDLINE=22483629; PubMed=12552318;
RA   Kayashima T., Yamasaki K., Yamada T., Sakai H., Miwa N., Ohta T.,
RA   Yoshiura K.-I., Matsumoto N., Nakane Y., Kanetake H., Ishino F.,
RA   Niikawa N., Kishino T.;
RT   "The novel imprinted carboxypeptidase A4 gene (CPA4) in the 7q32
RT   imprinting domain.";
RL   Hum. Genet. 112:220-226(2003).
CC   -!- FUNCTION: Could be involved in the histone hyperacetylation
CC       pathway.
CC   -!- TISSUE SPECIFICITY: Fetal expression in the adrenal gland, brain,
CC       heart, intestine, kidney, liver and lung. Except for fetal brain
CC       that shows no imprinting, expression was found preferentially from
CC       the maternal allele.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095719; AAF23230.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.017; -.
DR   Genew; HGNC:15740; CPA4.
DR   MIM; 607635; -.
DR   GO; GO:0004182; F:carboxypeptidase A activity; NAS.
DR   GO; GO:0016573; P:histone acetylation; NAS.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; NAS.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc; Zymogen; Signal.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17    113       ACTIVATION PEPTIDE (POTENTIAL).
FT   CHAIN       114    421       CARBOXYPEPTIDASE A4.
FT   METAL       181    181       ZINC (BY SIMILARITY).
FT   METAL       184    184       ZINC (BY SIMILARITY).
FT   METAL       308    308       ZINC (BY SIMILARITY).
FT   ACT_SITE    382    382       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    250    273       BY SIMILARITY.
SQ   SEQUENCE   421 AA;  47379 MW;  91883D25B705BC6A CRC64;

  Query Match           34.2%;  Score 458;  DB 1;  Length 421;
  Best Local Similarity   38.4%;  Pred. No. 5.2e-35;
  Matches  103;  Conservative   35;  Mismatches   82;  Indels   48;  Gaps    6;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||||  ||  :: |    ||:  :: || |||  |:|||| |  :   : |:|:: |
Db        120 YGAYHSLEAIYHEMDNIAADFPDLARRVKIGHSFENRPMYVLKFSTGKGVRRPAVWLNAG 179

Qy         66 IHAREWISPAFCLW----------------------FI-------------GHNRMWR 88
              ||:||||| |  :|                      |:             ||:||
Db        180 IHSREWISQATAIWTARKIVSDYQRDPAITSILEKMDIFLLPVANPDGYVYTQTQNRLWR 239

Qy         89 KNRSFYANNHCIGTDLNRN----FASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASF 144
              | ||    : ||| | |||    || |    ||| : ||| | | :  || |||:|  |
Db        240 KTRSRNPGSSCIGADPNRNWNASFAGK-----GASDNPCSEVYHGPHANSEVEVKSVVDF 294

Qy        145 LRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTH 204
              :::: |  | :| :||||| :::|| |:  |: | |||  ||  | :|:   |  | |
Db        295 IQKHGN-FKGFIDLHSYSQLLMYPYGYSVKKAPDAEELDKVARLAAKALASVS-GTEYQV 352

Qy        205 GHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |    |:| | |   || || |||::||
Db        353 GPTCTTVYPASGSSIDWAYDNGIKFAFT 380


RESULT 15
CBPZ_SIMVI
ID   CBPZ_SIMVI     STANDARD;      PRT;   304 AA.
AC   P42788;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Zinc carboxypeptidase (EC 3.4.17.-) (Fragment).
OS   Simulium vittatum (Black fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Nematocera; Chironomoidea;
OC   Simuliidae; Simulium.
OX   NCBI_TaxID=7192;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Gut;
RX   MEDLINE=94093864; PubMed=8269093;
RA   Ramos A., Mahowald A., Jacobs-Lorena M.;
RT   "Gut-specific genes from the black fly Simulium vittatum encoding
RT   trypsin-like and carboxypeptidase-like proteins.";
RL   Insect Mol. Biol. 1:149-163(1993).
CC   -!- FUNCTION: Involved in the digestion of the blood meal.
CC   -!- SUBCELLULAR LOCATION: Secreted.
```

```
CC   -!- TISSUE SPECIFICITY: Gut-specific.
CC   -!- INDUCTION: By blood meal.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M14.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08481; AAA18531.1; -.
DR   HSSP; P48052; 1AYE.
DR   MEROPS; M14.UPA; -.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Hydrolase; Carboxypeptidase; Metalloprotease; Zinc.
FT   NON_TER       1      1
FT   METAL        58     58       ZINC (BY SIMILARITY).
FT   METAL        61     61       ZINC (BY SIMILARITY).
FT   METAL       184    184       ZINC (BY SIMILARITY).
FT   ACT_SITE    236    236       PROTON DONOR (BY SIMILARITY).
FT   ACT_SITE    259    259       NUCLEOPHILE (BY SIMILARITY).
FT   DISULFID    125    148       BY SIMILARITY.
SQ   SEQUENCE   304 AA;  34849 MW;  2E6E3FF8A6AA9144 CRC64;

  Query Match           30.8%;  Score 411.5;  DB 1;  Length 304;
  Best Local Similarity   33.3%;  Pred. No. 7.2e-31;
  Matches   94;  Conservative   52;  Mismatches   85;  Indels   51;  Gaps    9;

Qy          8 QYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCGIH 67
              |||:| |||||:: : : ||: :  :  | |:|   :  :||| |:      : ::  ||
Db          1 QYHTLPEIYSWLDRLVQEHPEHVEPVVGGKSYEGREIRGVKVSYKK--GNPVVMVESNIH 58

Qy         68 AREWISPA---FCL-------------------WFI-------GH------NRMWRKNRS 92
              |||||: |   : |                   |:|       |:      :||||| ||
Db         59 AREWITAATTTYLLNELLTSKNSTIREMAENYDWYIFPVTNPDGYVYTHTTDRMWRKTRS 118

Qy         93 FYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQI 152
                 :: | ||| |||: : || |:| ||  |:||| |    || | :: : ||:    ||
Db        119 PNPDSLCAGTDPNRNW-NFHWMEQGTSSRPCTETYGGKKAFSEVETRSFSDFLKTLKGQI 177

Qy        153 KAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSETLY 212
              | |:: ||||| ::||| :|   : :|::|  :   | |:: :    | || |:  : :|
Db        178 KVYLAFHSYSQLLLFPYGHTCQHTYNHDDLQAIGDAAARSLAQ-RYGTDYTVGNIYDAIY 236

Qy        213 LAPGGGDDWIY---DLGIKYSFTSNP---------PVEKLLP 242
               | ||  || |   |: | |::   |         |  :::|
Db        237 PASGGSMDWAYDTLDIPIAYTYELRPRDGWNGFQLPANQIIP 278


Search completed: January  1, 2004, 19:29:01
Job time : 10.7578 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 1, 2004, 19:19:56 ; Search time 41.5992 Seconds
(without alignments)
1526.015 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1327	99.2	360	4	Q9P2Y6	Q9p2y6 homo sapien
2	1239.5	92.6	423	4	Q15114	Q15114 homo sapien
3	1234.5	92.3	423	4	Q96IY4	Q96iy4 homo sapien
4	1105.5	82.6	422	11	Q9JHH6	Q9jhh6 mus musculu
5	1095.5	81.9	422	11	Q9EQV9	Q9eqv9 rattus norv
6	1095.5	81.9	422	11	Q9QZF0	Q9qzf0 mus musculu
7	622.5	46.5	408	13	Q8AXN3	Q8axn3 paralichthy
8	612.5	45.8	416	13	Q9PUF2	Q9puf2 bothrops ja
9	582	43.5	416	6	Q9XSP3	Q9xsp3 sus scrofa
10	557	41.6	412	11	P97597	P97597 rattus norv
11	550.5	41.1	374	4	Q8IVL8	Q8ivl8 homo sapien
12	547	40.9	437	4	Q8N4T0	Q8n4t0 homo sapien
13	544.5	40.7	434	13	Q8AVS2	Q8avs2 xenopus lae
14	514.5	38.5	419	13	Q8UUK1	Q8uuk1 gallus gall
15	488.5	36.5	419	6	Q9TV85	Q9tv85 sus scrofa
16	486.5	36.4	436	11	Q8R4H4	Q8r4h4 mus musculu
17	484.5	36.2	262	11	Q8BVD0	Q8bvd0 mus musculu
18	477.5	35.7	247	4	Q96KZ9	Q96kz9 homo sapien
19	476	35.6	420	11	Q8BMK6	Q8bmk6 mus musculu
20	465.5	34.8	436	4	Q8WXQ8	Q8wxq8 homo sapien
21	456.5	34.1	419	13	Q8AXN4	Q8axn4 paralichthy
22	456	34.1	198	4	Q9NTI8	Q9nti8 homo sapien
23	449.5	33.6	226	4	Q8TDE8	Q8tde8 homo sapien
24	428	32.0	452	13	Q8UW45	Q8uw45 fugu rubrip
25	426	31.8	1132	5	Q9W475	Q9w475 drosophila
26	424	31.7	424	5	Q9VL87	Q9vl87 drosophila
27	424	31.7	428	5	Q8IGC4	Q8igc4 drosophila
28	419	31.3	351	5	Q8T0Q1	Q8t0q1 drosophila
29	419	31.3	422	5	Q9VLZ2	Q9vlz2 drosophila
30	419	31.3	445	5	Q9W478	Q9w478 drosophila
31	406.5	30.4	419	13	Q8AXN5	Q8axn5 paralichthy
32	403	30.1	323	5	Q19121	Q19121 caenorhabdi
33	402	30.0	416	5	Q9VRZ3	Q9vrz3 drosophila
34	396.5	29.6	584	5	Q9TZH1	Q9tzh1 caenorhabdi
35	385.5	28.8	453	5	Q9V342	Q9v342 drosophila
36	385.5	28.8	454	5	Q9TZC6	Q9tzc6 caenorhabdi
37	382.5	28.6	341	5	Q8SXR8	Q8sxr8 drosophila
38	382.5	28.6	424	5	Q9VS66	Q9vs66 drosophila
39	380	28.4	430	5	Q9VL86	Q9vl86 drosophila
40	379	28.3	351	4	Q8NA08	Q8na08 homo sapien
41	377.5	28.2	423	5	O61532	O61532 drosophila
42	375	28.0	427	5	Q9U9K2	Q9u9k2 aedes aegyp
43	374.5	28.0	488	5	Q9BL88	Q9bl88 caenorhabdi
44	373	27.9	430	5	Q8SZP6	Q8szp6 drosophila
45	370.5	27.7	419	5	Q9VLZ1	Q9vlz1 drosophila

ALIGNMENTS

RESULT 1
Q9P2Y6
ID Q9P2Y6 PRELIMINARY; PRT; 360 AA.
AC Q9P2Y6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Carboxypeptidase B-like protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matsumoto A.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92042093; PubMed=1939207;
RA Eaton D.L., Malloy B.E., Tsai S., Henzel W., Drayna D.;
RT "Isolation, molecular cloning, and partial characterization of a novel
RT carboxypeptidase B from human plasma.";
RL J. Biol. Chem. 266:21833-21838(1991).
DR EMBL; AB011969; BAA90475.1; -.
DR HSSP; P00730; 2CTC.
DR MEROPS; M14.009; -.
DR InterPro; IPR003146; Propep_M14.
DR InterPro; IPR000834; Zn_carbOpept.
DR Pfam; PF02244; Propep_M14; 1.
DR Pfam; PF00246; Zn_carbOpept; 1.
DR PRINTS; PR00765; CRBOXYPTASEA.
DR SMART; SM00631; Zn_pept; 1.
SQ SEQUENCE 360 AA; 40935 MW; BF670B2F7437C1CB CRC64;

Query Match 99.2%; Score 1327; DB 4; Length 360;
Best Local Similarity 99.2%; Pred. No. 1e-119;
Matches 244; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASS 234

Qy        121 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 SSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHE 294

Qy        181 ELSLVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 240
              ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        295 ELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTSNPPVEKL 354

Qy        241 LPLSLK 246
              ||||||
Db        355 LPLSLK 360


RESULT 2
Q15114
ID   Q15114      PRELIMINARY;      PRT;   423 AA.
AC   Q15114;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PCPB protein.
GN   PCPB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=92042093; PubMed=1939207;
RA   Eaton D.L., Malloy B.E., Tsai S.P., Henzel W., Drayna D.;
RT   "Isolation, molecular cloning, and partial characterization of a novel
RT   carboxypeptidase B from human plasma.";
RL   J. Biol. Chem. 266:21833-21838(1991).
DR   EMBL; M75106; AAA60042.1; -.
DR   HSSP; P00730; 1PYT.
DR   MEROPS; M14.009; -.
DR   Genew; HGNC:2300; CPB2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
FT   CHAIN       23    423       PLASMA CARBOXYPEPTIDASE B.
SQ   SEQUENCE   423 AA;  48442 MW;  9B383272F6EE79F4 CRC64;

  Query Match             92.6%;  Score 1239.5;  DB 4;  Length 423;
  Best Local Similarity   86.2%;  Pred. No. 3.6e-111;
  Matches  232;  Conservative    0;  Mismatches    0;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH----------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 3
Q96IY4
ID   Q96IY4      PRELIMINARY;      PRT;   423 AA.
AC   Q96IY4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B2 (Plasma).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skeletal muscle;
RA   Strausberg R.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC007057; AAH07057.1; -.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   423 AA;  48412 MW;  9B383E03F6EE7CF5 CRC64;

  Query Match             92.3%;  Score 1234.5;  DB 4;  Length 423;
  Best Local Similarity   85.9%;  Pred. No. 1.1e-110;
  Matches  231;  Conservative    0;  Mismatches    1;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db        115 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQAAKNAI 174

Qy         61 WIDCGIHAREWISPAFCLWFIGH----------------------------------- 83
              |||||||||||||||||||||||
Db        175 WIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFYVMPVVNVDGYDYSWKK 234

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 294

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        295 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 354

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |||||||||||||||||||||||||||||
Db        355 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 383


RESULT 4
Q9JHH6
ID   Q9JHH6      PRELIMINARY;      PRT;   422 AA.
AC   Q9JHH6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase R (Thrombin-activatable fibrinolysis inhibitor)
DE   (1110032P04Rik protein).
GN   CPB2 OR TAFI OR 1110032P04RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20341711; PubMed=10878383;
RA   Sato T., Miwa T., Akatsu H., Matsukawa N., Obata K., Okada N.,
RA   Campbell W., Okada H.;
RT   "Pro-carboxypeptidase R is an acute phase protein in the mouse,
RT   whereas carboxypeptidase N is not.";
RL   J. Immunol. 165:1053-1058(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20201996; PubMed=10739389;
RA   Marx P.F., Wagenaar G.T.M., Reijerkerk A., Tiekstra M.J.,
RA   van Rossum A.G.S.H., Gebbink M.F.G.B., Meijers J.C.M.;
RT   "Characterization of mouse thrombin-activatable fibrinolysis
RT   inhibitor.";
RL   Thromb. Haemost. 83:297-303(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Embryo;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AB021968; BAB03402.1; -.
DR   EMBL; AF164524; AAF62385.1; -.
DR   EMBL; AK004045; BAB23141.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48870 MW;  99113755669D55CB CRC64;

  Query Match            82.6%;  Score 1105.5;  DB 11;  Length 422;
  Best Local Similarity  77.0%;  Pred. No. 3.1e-98;
  Matches  207;  Conservative   10;  Mismatches   15;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||| |||:||||| ||:||||||||||||||||||  ||||
Db        114 ASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAI 173

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||:
Db        174 WIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYIMPVMNVDGYDYTWKK 233

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||| : || |:||||||||||||||:||||||||||||||||||||||||||||
Db        234 NRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSETYCGLYPESEPEVKAVAD 293

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||: |||||||||||| |:||||| ||||||||||||||||||||  :||||||
Db        294 FLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYT 353

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              || |||:||||||| ||||||||||||||
Db        354 HGSGSESLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 5
Q9EQV9
ID   Q9EQV9      PRELIMINARY;      PRT;   422 AA.
AC   Q9EQV9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Pre-procarboxypeptidase R.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20471387; PubMed=11021404;
RA   Kato T., Sato T., Matsuo S., Yamamoto T., Campbell W., Hotta N.,
RA   Okada N., Okada H.;
RT   "Molecular cloning and partial characterization of rat
RT   procarboxypeptidase R and carboxypeptidase N.";
RL   Microbiol. Immunol. 44:719-728(2000).
DR   EMBL; AB042598; BAB18617.1; -.
DR   HSSP; P00730; 2CTC.
DR   MEROPS; M14.009; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   SMART; SM00631; Zn_pept; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   422 AA;  48826 MW;  FFFD32A51A9366C8 CRC64;

  Query Match            81.9%;  Score 1095.5;  DB 11;  Length 422;
  Best Local Similarity  75.8%;  Pred. No. 2.8e-97;
  Matches  204;  Conservative   11;  Mismatches   17;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              ||:|||||||||||||||| |||:||||| ||:||||:||||||||||||   ||||
Db        114 ASSSYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSYEKYPLYVLKVSGKEHRVKNAI 173

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||:
Db        174 WIDCGIHAREWISPAFCLWFIGYVTQFHGKENTYTRLLRHVDFYIMPVMNVDGYDYTWKK 233

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||| : || |:||||||||||||||:|||| ||||||||||||||||||||||
Db        234 NRMWRKNRSVHMNNRCVGTDLNRNFASKHWCEKGASSFSCSETYCGLYPESEPEVKAVAD 293

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||| |||||||||||| |:||||| ||||||||||||||||||||  :||||||
Db        294 FLRRNINHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIESINKNTRYT 353

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              || |||:||||||| ||||||||||||||
Db        354 HGSGSESLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 6
Q9QZF0
ID   Q9QZF0      PRELIMINARY;      PRT;   422 AA.
AC   Q9QZF0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase U.
GN   CPB2.
OS   Mus musculus (Mouse).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B6/CBAF1J;
RA   He Y.C., Broze G.;
RT   "Isolation and characterization of mouse liver carboxypeptidase B
RT   gene.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF186188; AAF00528.1; -.
DR   HSSP; P00730; 5CPA.
DR   MEROPS; M14.009; -.
DR   MGD; MGI:1891837; Cpb2.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
SQ   SEQUENCE   422 AA;  48832 MW;  B53FFFB09943954E CRC64;

  Query Match             81.9%;  Score 1095.5;  DB 11;  Length 422;
  Best Local Similarity   76.2%;  Pred. No. 2.8e-97;
  Matches  205;  Conservative   11;  Mismatches   16;  Indels   37;  Gaps    1;

Qy          1 ASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              |||||||||||||||||||| :||:||||| ||:||||||||||||||||||||  ||||
Db        114 ASASYYEQYHSLNEIYSWIEVLTEQHPDMLQKIYIGSSFEKYPLYVLKVSGKEQRIKNAI 173

Qy         61 WIDCGIHAREWISPAFCLWFIGH---------------------------------- 83
              |||||||||||||||||||||:
Db        174 WIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYIMPVMNVDGYDYTWKK 233

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              ||||||||| : || |:||||||||||||||:|||||||||||||||||||||||||||
Db        234 NRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSETYCGLYPESEPEVKAVAD 293

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              ||||||: |||||||||||| |:||||| ||||||||||||||||||   :||||||
Db        294 FLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSLVASEAVRAIGSFNKNTRYT 353

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              || |||:||||||| ||||||||||||||
Db        354 HGSGSESLYLAPGGSDDWIYDLGIKYSFT 382


RESULT 7
Q8AXN3
ID   Q8AXN3      PRELIMINARY;      PRT;   408 AA.
AC   Q8AXN3;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase B (Fragment).
GN   CPB.
OS   Paralichthys olivaceus (Flounder).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC   Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC   Pleuronectoidei; Paralichthyidae; Paralichthys.
OX   NCBI_TaxID=8255;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Srivastava A.S., Kurokawa T., Suzuki T.;
RT   "Molecular cloning and cDNA sequence analysis of carboxypeptidase A1,
RT   A2 and B from the Japanese flounder, Paralichthys olivaceus.";
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB099302; BAC53789.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   408 AA;  45904 MW;  809CAACD9D2DA34B CRC64;

  Query Match             46.5%;  Score 622.5;  DB 13;  Length 408;
  Best Local Similarity   46.1%;  Pred. No. 1.2e-51;
  Matches  123;  Conservative   41;  Mismatches   64;  Indels   39;  Gaps    4;

Qy          3 ASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWI 62
              | | :|::|::: ||| |:  :||:::|  ||:::|  |: |||:  |  : | ||::
Db        103 AHSYTKYNNLDKVQSWIASISSSNPDLISKQVIGNTYEGRPMTVLKLGKKSSSTKPAIFM 162

Qy         63 DCGIHAREWISPAFCLWFI-----------------------------GH------NR 85
              |||||||||||||| ||:                             |:      ||
Db        163 DCGIHAREWISPAFCQWFVKEALSTYGSDSQMTSLLNQMDVFVLPVFNIDGYDFTHKSNR 222

Qy         86 MWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFL 145
              |||| ||  : : ||| | ||||    ||  ||| : ||:|:||  |||| ||| || |:
Db        223 MWRKTRSRKSGSSCIGADPNRNF-DAGWCSMGASDNPCSDTFCGYTPESEIEVKNVADFI 281

Qy        146 RRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHG 205
              ||| : ||||:::|||||| ::||||||   : || ||  ||  |  |: ::   | || |
Db        282 RRNRSIIKAYLTVHSYSQLLLFPYSYTYQLAADHSELMKVAEGASAAL-RSLYGTTYTSG 340

Qy        206 HGSETLYLAPGGGDDWIYDLGIKYSFT 232
              |: |:| | || ||| ||||:|||:|
Db        341 PGATTIYPAAGGSDDWAYDLGVKYSYT 367


RESULT 8
Q9PUF2
ID   Q9PUF2      PRELIMINARY;      PRT;   416 AA.
AC   Q9PUF2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase homolog.
OS   Bothrops jararaca (Jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
OX   NCBI_TaxID=8724;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RT   "Screening of Bothrops jararaca pancreas cDNA library.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF190274; AAF01344.1; -.
DR   HSSP; P09955; 1NSA.
DR   MEROPS; M14.003; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
SQ   SEQUENCE   416 AA;  47723 MW;  4F999854DD72B7A7 CRC64;

  Query Match             45.8%;  Score 612.5;  DB 13;  Length 416;
  Best Local Similarity   45.0%;  Pred. No. 1.2e-50;
  Matches  121;  Conservative   44;  Mismatches   63;  Indels   41;  Gaps    5;

Qy          2 SASY-YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAI 60
              :| | ||:|:|  :| :|   |   :| ::::: ||::||  |: :||| ||    | ||
Db        111 TAGYNYEKYNSWEKIDAWTADIANENPSLVSRLQIGTTFEGRPMPLLKV-GKPGVNKKAI 169

Qy         61 WIDCGIHAREWISPAFCLWFI-----------------------------------GH 83
              :|||| ||||||||||| ||:
Db        170 FIDCGFHAREWISPAFCQWFVREAVRTYGKETIMTQLLNKLDFYILPVLNIDGYVYSWKQ 229

Qy         84 NRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVAS 143
              :||||| ||  | : ||||| |||| :  ||  ||| : |||||||  |||| | ||:|
Db        230 SRMWRKTRSVNAGSTCIGTDPNRNFDAA-WCSVGASRNPCSETYCGSKPESEKETKALAD 288

Qy        144 FLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYT 203
              |:||| : |:||:::||||| :::|||||   : ::::|: :| ||:| : |    | ||
```

```
Db        289 FIRRNRSIIQAYLTIHSYSQMLLYPYSYTYDLTSNNKKLNSIAKEAIREL-KVLFGTEYT 347

Qy        204 HGHGSETLYLAPGGGDDWIYDLGIKYSFT 232
              :| |: |:| | || ||| || ||||:||
Db        348 YGPGAATIYPAAGGSDDWAYDQGIKYAFT 376


RESULT 9
Q9XSP3
ID   Q9XSP3      PRELIMINARY;      PRT;   416 AA.
AC   Q9XSP3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Procarboxypeptidase B precursor (EC 3.4.17.2).
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RX   MEDLINE=99321926; PubMed=10391940;
RA   Ventura S., Villegas V., Sterner J., Larson J., Vendrell J.,
RA   Hershberger C., Aviles F.;
RT   "Mapping the Pro-region of carboxypeptidase B by protein engineering.
RT   Cloning, overexpression, and mutagenesis of the porcine proenzyme.";
RL   J. Biol. Chem. 274:19925-19933(1999).
DR   EMBL; AJ133775; CAB46991.1; -.
DR   HSSP; P09955; 1PBA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Hydrolase; Signal.
FT   SIGNAL       1     15       POTENTIAL.
FT   CHAIN       16    416       PROCARBOXYPEPTIDASE B.
SQ   SEQUENCE   416 AA;  47381 MW;  5FA7FCED7B45AE6A CRC64;

  Query Match             43.5%;  Score 582;  DB 6;  Length 416;
  Best Local Similarity   44.3%;  Pred. No. 1e-47;
  Matches  117;  Conservative   40;  Mismatches   67;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              ||:|::   | :| : :|  :||::::  ||::|    :|:||| ||     | ||::|||
Db        116 YEKYNNWETIEAWTKQVTSENPDLISRTAIGTTFLGNNIYLLKV-GKPGPNKPAIFMDCG 174

Qy         66 IHAREWISPAFCLWFI-------------------------------GH------NRMWR 88
               ||||||| ||| ||:                               |:      |||||
Db        175 FHAREWISHAFCQWFVREAVLTYGYESHMTEFLNKLDFYVLPVLNIDGYIYTWTKNRMWR 234

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  |   ||||| ||||    ||  |||:  | |||||   ||| | ||:| |:| |
Db        235 KTRSTNAGTTCIGTDPNRNF-DAGWCTTGASTDPCDETYCGSAAESEKETKALADFIRNN 293

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :: ||||:::|||| |::|||      ::: ||: :|  ||: :  |   |:||:| |:
Db        294 LSSIKAYLTIHSYSQMILYPYSYDYKLPENNAELNNLAKAAVKEL-ATLYGTKYTYGPGA 352

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              |:| | || ||| || |||||||
Db        353 TTIYPAAGGSDDWAYDQGIKYSFT 376


RESULT 10
P97597
ID   P97597      PRELIMINARY;      PRT;   412 AA.
AC   P97597;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Mast cell carboxypeptidase A precursor (Fragment).
GN   R-CPA.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague Dawley;
RX   MEDLINE=97149430; PubMed=8996238;
RA   Lutzelschwab C., Pejler G., Aveskogh M., Hellman L.;
RT   "Secretory granule proteases in rat mast cells. Cloning of 10
RT   different serine proteases and a carboxypeptidase A from various rat
RT   mast cell populations.";
RL   J. Exp. Med. 185:13-29(1997).
DR   EMBL; U67914; AAB48267.1; -.
DR   HSSP; P09955; 1NSA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Signal.
FT   NON_TER      1      1
FT   SIGNAL      <1     10       POTENTIAL.
FT   CHAIN       11    412       MAST CELL CARBOXYPEPTIDASE A.
SQ   SEQUENCE   412 AA;  47944 MW;  DAB59555FC49137D CRC64;

  Query Match             41.6%;  Score 557;  DB 11;  Length 412;
  Best Local Similarity   43.0%;  Pred. No. 2.6e-45;
  Matches  113;  Conservative   38;  Mismatches   72;  Indels   40;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|:  |:| || | : |:||:|:::| |||: |  ||||||: |::   : ||::|||
Db        111 YAKYNDWNKIVSWTEKMVEKHPEMVSRIKIGSTVEDNPLYVLKI-GRKDGERKAIFMDCG 169

Qy         66 IHAREWISPAFCLWFI-------GHN-----------------------------RMWR 88
              ||||||:||||| ||:       | |                             ||||
Db        170 IHAREWVSPAFCQWFVYQAAKSYGKNKIMTKLLDRMNFYVLPVFNVDGYIWSWTKDRMWR 229

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              ||||   |: ||||||||||    |     : : |   | |  |||| | ||| :|:| :
Db        230 KNRSKNPNSTCIGTDLNRNF-DVSWDSSPNTDNPCLSVYRGPAPESEKETKAVTNFIRSH 288

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
              :| |||||: ||||| ::||| ||      :|::|  ||  |   :  :   ||| :|  :
Db        289 LNSIKAYITFHSYSQMLLFPYGYTIKLPPNHQDLLKVARIATDVL-SSRYETRYIYGPIA 347

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              |:|   |   || |||||::|
Db        348 STIYKTSGSSLDWAYDLGIKHTF 370


RESULT 11
Q8IVL8
ID   Q8IVL8      PRELIMINARY;      PRT;   374 AA.
AC   Q8IVL8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Zn-carboxypeptidase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
```

```
RA   Obaya A.J., Lopez-Otin C.;
RT   "A new Zn-carboxypeptidase highly expressed in ovary.";
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ422118; CAD19478.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   374 AA;  42529 MW;  404C373BB841AAD2 CRC64;

  Query Match             41.1%;  Score 550.5;  DB 4;  Length 374;
  Best Local Similarity   39.4%;  Pred. No. 9.7e-45;
  Matches  104;  Conservative   45;  Mismatches   76;  Indels   39;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  || : ||| |:  |:|:: :::|:  :| ::| :|:| ||:|      |  ||:|||
Db         47 YNIYHPMGEIYEWMREISEKYKEVVTQHFLGVTYETHPMYYLKISQPSGNPKKIIWMDCG 106

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              |||||||:|||| ||:                              |:      :|:||
Db        107 IHAREWIAPAFCQWFVKEILQNHKDNSSIRKLLRNLDFYVLPVLNIDGYIYTWTTDRLWR 166

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              |:|| : |  | |||||||| :  ||  ||| :  :|:||  | |||| ||||||:
Db        167 KSRSPHNNGTCFGTDLNRNF-NASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIESK 225

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
               : |  :::|||| | |: || ||::|| :| |:  |  :|  |: |    | |  |  :
Db        226 KDDILCFLTMHSYGQLILTPYGYTKNKSSNHPEMIQVGQKAANAL-KAKYGTNYRVGSSA 284

Qy        209 ETLYLAPGGGDDWIYDLGIKYSFT 232
              : || : |   ||  |:|| :|:|
Db        285 DILYASSGSSRDWARDIGIPFSYT 308


RESULT 12
Q8N4T0
ID   Q8N4T0      PRELIMINARY;      PRT;   437 AA.
AC   Q8N4T0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC033684; AAH33684.1; -.
DR   Genew; HGNC:17245; CPA6.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase.
SQ   SEQUENCE   437 AA;  51024 MW;  9970680D429A0A98 CRC64;

  Query Match             40.9%;  Score 547;  DB 4;  Length 437;
  Best Local Similarity   40.3%;  Pred. No. 2.6e-44;
  Matches  106;  Conservative   40;  Mismatches   77;  Indels   40;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              || |||| || :|:  : : |  ::    || |:|   |::|: |:    | |:|||||
Db        136 YEVYHSLEEIQNWMHHLNKTHSGLIHMFSIGRSYEGRCLFILKL-GRRSRLKRAVWIDCG 194

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWR 88
              ||||||| |||| ||:                                   ::| ||
Db        195 IHAREWIGPAFCQWFVKEALLTYKSDPAMRKMLNHLYFYIMPVFNVDGYHFSWTNDRFWR 254

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||  :   | | | |||:  | ||:||||   | :|||| :||||||||||:|||::
Db        255 KTRSRNSRFRCRGVDANRNWKVK-WCDEGASMHPCDDTYCGPFPESEPEVKAVANFLRKH 313

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
                 |:||:| |:|:| :::|||| :   :   :   | :|| |::      || :|  |
Db        314 RKHIRAYLSFHAYAQMLLYPYSYKYATIPNFRCVESAAYKAVNALQSV-YGVRYRYGPAS 372

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              |||:: |   || |  || |:|
Db        373 TTLYVSSGSSMDWAYKNGIPYAF 395


RESULT 13
Q8AVS2
ID   Q8AVS2      PRELIMINARY;      PRT;   434 AA.
AC   Q8AVS2;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to carboxypeptidase A6.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RA   Klein S., Strausberg R.;
RL   Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC041315; AAH41315.1; -.
KW   Carboxypeptidase.
SQ   SEQUENCE   434 AA;  50581 MW;  B42BD77843910591 CRC64;

  Query Match             40.7%;  Score 544.5;  DB 13;  Length 434;
  Best Local Similarity   41.1%;  Pred. No. 4.5e-44;
  Matches  108;  Conservative   36;  Mismatches   80;  Indels   39;  Gaps    4;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              | :|| |:|| ||: :: : | |:::   || |:|   |||||:    :: | |||||||
Db        132 YNEYHPLHEIESWMFYMNKTHHDLVSLFTIGKSYEGRSLYVLKLGKDTKSYKKAIWIDCG 191

Qy         66 IHAREWISPAFCLWFI------------------------------GH------NRMWR 88
              :||||| |||| ||:                              |:      :| ||
Db        192 MHAREWIGPAFCQWFVKEAINSYNTDPAMKKILNLLYIYVMPVFNVDGYHYSWHSDRFWR 251

Qy         89 KNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRN 148
              | ||      | | | |||:   || :|||| : |  |||| | |||||||||| || :
Db        252 KTRSKNTRYQCYGVDANRNW-KVHWSDEGASLNPCDNTYCGPYAESEPEVKAVAQFLYKQ 310

Qy        149 INQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGS 208
                 ::||:| |:|:| :::|||| :   :   |  || || :::   || ||  |
Db        311 RKHVRAYMSFHAYAQMLLYPYSYQYGAIPNFGCVESAAHNAVLAI-RSAYGIRYRHGPAS 369

Qy        209 ETLYLAPGGGDDWIYDLGIKYSF 231
              ||||  |   || |: || ||:
Db        370 STLYLTSGSSMDWAYNNGIPYSY 392


RESULT 14
Q8UUK1
ID   Q8UUK1      PRELIMINARY;      PRT;   419 AA.
AC   Q8UUK1;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Preprocarboxypeptidase A precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=domesticus; TISSUE=Pancreas;
RA   Hasegawa S., Honda K., Hikami Y.;
RT   "Nucleotide sequence of cDNA encoding chicken carboxypeptidase A
RT   precursor.";
RL   Submitted (FEB-1992) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X64539; CAA45837.1; -.
DR   MEROPS; M14.001; -.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Signal; Carboxypeptidase.
FT   SIGNAL       1     17       POTENTIAL.
FT   CHAIN      113    419       CARBOXYPEPTIDASE A.
SQ   SEQUENCE   419 AA;  47006 MW;  299C334274BCBC9C CRC64;

  Query Match             38.5%;  Score 514.5;  DB 13;  Length 419;
  Best Local Similarity   42.0%;  Pred. No. 3.4e-41;
  Matches  110;  Conservative   40;  Mismatches   73;  Indels   39;  Gaps    5;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKNAIWIDCG 65
              |  ||:| |||:::: :   :|::::|: || : |  |:|||| | |  | : ||||| |
Db        121 YTSYHTLEEIYAFMDLLVAENPNLVSKLEIGRTTENRPIYVLKFS-KGGTNRPAIWIDTG 179

Qy         66 IHAREWISPAFCLWFI-----------------------------------GHNRMWRKN 90
              ||:|||:: |  :||                                     ||||||
Db        180 IHSREWVTQASGIWFAKKIVQEQDEGLANILDQMDIFLEIVTNPDGFAFTHTSNRMWRKT 239

Qy         91 RSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNIN 150
              ||  : : |:| | |||:    :   ||||: ||||| | |  |||||||:  |:: : |
Db        240 RSKRSGSLCVGVDPNRNW-DAGFGGSGASSNPCSETYHGPYANSEPEVKAIVDFVKSHKN 298

Qy        151 QIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHGHGSET 210
               |||:||:||||| :::|| || :   | |||  || ||| |:  :   | | :|    |
Db        299 -IKAFISIHSYSQLLLYPYGYTTTAVPDKEELHQVAKEAVAAL-SSLYGTNYKYGSIITT 356

Qy        211 LYLAPGGGDDWIYDLGIKYSFT 232
              :| | ||  || |: |||||||
Db        357 IYQASGGTIDWTYNQGIKYSFT 378


RESULT 15
Q9TV85
ID   Q9TV85      PRELIMINARY;      PRT;   419 AA.
AC   Q9TV85;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Carboxypeptidase A1 (EC 3.4.17.1).
GN   PCPA1.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99192816; PubMed=10092856;
RA   Darnis S., Juge N., Marino C., Aviles F.X., Puigserver A., Chaix J.C.,
RA   Guo X.J.;
RT   "Cloning, sequencing and functional expression of a cDNA encoding
RT   porcine pancreatic preprocarboxypeptidase A1.";
RL   Eur. J. Biochem. 259:719-725(1999).
DR   EMBL; AF076222; AAD17690.1; -.
DR   HSSP; P09954; 1PCA.
DR   InterPro; IPR003146; Propep_M14.
DR   InterPro; IPR000834; Zn_carbOpept.
DR   Pfam; PF02244; Propep_M14; 1.
DR   Pfam; PF00246; Zn_carbOpept; 1.
DR   PRINTS; PR00765; CRBOXYPTASEA.
DR   SMART; SM00631; Zn_pept; 1.
DR   PROSITE; PS00132; CARBOXYPEPT_ZN_1; 1.
DR   PROSITE; PS00133; CARBOXYPEPT_ZN_2; 1.
KW   Carboxypeptidase; Hydrolase.
SQ   SEQUENCE   419 AA;  47235 MW;  84B4CB557B714FC1 CRC64;

  Query Match             36.5%;  Score 488.5;  DB 6;  Length 419;
  Best Local Similarity   39.0%;  Pred. No. 1.1e-38;
  Matches  104;  Conservative   42;  Mismatches   74;  Indels   47;  Gaps    6;

Qy          6 YEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYVLKVSGKEQTAKN---AIWI 62
              |  ||:| ||| ::: :   || :::|: ||||:|  |:|||| |    |  |   ||||
Db        119 YATYHTLEEIYDFMDILVAEHPQLVSKLQIGSSYEGRPIYVLKFS----TGGNNRPAIWI 174

Qy         63 DCGIHAREWISPAFCLWFI-----------------------------------GHNR 85
              | |||:|||:: |  :||                                     ||
Db        175 DTGIHSREWVTQASGVWFAKKITEDYGQDPAFTAILDNLDIFLEIVTNPDGFAFTHSENR 234

Qy         86 MWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFL 145
              |||| ||  : : |:| | |||:    :   ||||: ||||| | :| || |||::  |:
Db        235 MWRKTRSRTSGSFCVGVDPNRNW-DAGFGGAGASSNPCSETYHGKFPNSEVEVKSIVDFV 293

Qy        146 RRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKTSKNTRYTHG 205
               : | |||:||:||||| :::|| |       | :||  ::  || |:  :   |:: :|
Db        294 NDHGN-IKAFISIHSYSQLLLYPYGYKTEAPADKDELDQISKSAVAAL-TSLYGTKFQYG 351

Qy        206 HGSETLYLAPGGGDDWIYDLGIKYSFT 232
                 |:| | ||  || |: ||||||:
Db        352 SIITTIYQASGGTIDWTYNQGIKYSFS 378


Search completed: January  1, 2004, 19:31:58
Job time : 43.5992 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 1, 2004, 19:34:47 ; Search time 3798.88 Seconds
(without alignments)
2649.143 Million cell updates/sec

Title: US-09-980-881A-4
Perfect score: 1338
Sequence: 1 ASASYYEQYHSLNEIYSWIE.....IKYSFTSNPPVEKLLPLSLK 246

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09980881/runat_31122003_135915_26829/app_query.fasta_1.1628
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980881_@CGN_1_1_8325_@runat_31122003_135915_26829 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1327	99.2	1546	9	AB011969	AB011969 Homo sapi
2	1289.5	96.4	1625	6	AR232183	AR232183 Sequence
3	1272	95.1	1272	6	AR086324	AR086324 Sequence
4	1272	95.1	1272	6	BD106653	BD106653 Method of
5	1272	95.1	1728	6	AX409750	AX409750 Sequence
6	1272	95.1	1728	6	AX706330	AX706330 Sequence
7	1272	95.1	1728	9	HUMPCPBX	M75106 Human prepr
8	1272	95.1	1749	6	I16100	I16100 Sequence 2
9	1272	95.1	1749	6	I33526	I33526 Sequence 2
10	1267	94.7	1272	9	BT006936	BT006936 Homo sapi
11	1267	94.7	1272	12	BT007990	BT007990 Synthetic
12	1267	94.7	1715	9	BC007057	BC007057 Homo sapi
13	1112	83.1	1421	10	AB021968	AB021968 Mus muscu
14	1112	83.1	1490	10	AF164524	AF164524 Mus muscu
15	1102	82.4	1430	10	AF186188	AF186188 Mus muscu
16	1099	82.1	1425	10	AB042598	AB042598 Rattus no
17	622.5	46.5	1441	5	AB099302	AB099302 Paralicht
18	612.5	45.8	1327	5	AF190274	AF190274 Bothrops
19	610	45.6	927	6	AR072911	AR072911 Sequence
20	582	43.5	921	6	I67699	I67699 Sequence 3
21	582	43.5	1215	6	I67698	I67698 Sequence 1
22	582	43.5	1251	4	SSC133775	AJ133775 Sus scrof
23	578	43.2	1251	4	DOGZAP47	D78348 Canis famil
24	567	42.4	999	6	A51908	A51908 Sequence 72
25	567	42.4	999	6	AR085871	AR085871 Sequence
26	567	42.4	1053	6	A51904	A51904 Sequence 68
27	567	42.4	1053	6	AR085868	AR085868 Sequence
28	567	42.4	1254	12	BT008111	BT008111 Synthetic
29	567	42.4	1263	6	A51896	A51896 Sequence 60
30	567	42.4	1263	6	A87530	A87530 Sequence 11
31	567	42.4	1263	6	AR085861	AR085861 Sequence
32	567	42.4	1284	6	A51913	A51913 Sequence 77
33	567	42.4	1284	6	AR085875	AR085875 Sequence
34	567	42.4	1310	9	HSA224866	AJ224866 Homo sapi
35	567	42.4	1454	9	BC015338	BC015338 Homo sapi
36	561	41.9	1442	10	MUSCARA	J05118 Mouse mast
37	560	41.9	1053	6	AR300428	AR300428 Sequence
c 38	560	41.9	1053	6	AR300429	AR300429 Sequence
39	560	41.9	1053	6	AX000417	AX000417 Sequence
c 40	560	41.9	1053	6	AX000418	AX000418 Sequence
41	560	41.9	1053	6	BD080205	BD080205 Expressio
c 42	560	41.9	1053	6	BD080206	BD080206 Expressio
43	560	41.9	1059	6	A51917	A51917 Sequence 81
44	560	41.9	1059	6	AR085878	AR085878 Sequence
45	559	41.8	1059	6	A51921	A51921 Sequence 85

ALIGNMENTS

RESULT 1

```
AB011969
LOCUS       AB011969               1546 bp    mRNA    linear   PRI 02-FEB-2000
DEFINITION  Homo sapiens mRNA for carboxypeptidase B-like protein, complete
            cds.
ACCESSION   AB011969
VERSION     AB011969.1  GI:6855463
KEYWORDS    carboxypeptidase B-like protein.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
REFERENCE   2  (sites)
  AUTHORS   Matsumoto,A.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 1546)
  AUTHORS   Matsumoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1998) Akira Matsumoto, Kobe University School of
            Medicine, Department of Radiation Biophysics & Genetics;
            Kusunoki-cho 7-5-1, Chuo-ku, Kobe, Hyogo 650-0017, Japan
            (E-mail:amat@med.kobe-u.ac.jp, Tel:81-78341-7451)
COMMENT     Sequence updated (10-Apr-1998).
FEATURES             Location/Qualifiers
     source          1. .1546
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     CDS             20. .1102
                     /note="carboxypeptidase B-like protein in human brain"
                     /codon_start=1
                     /product="carboxypeptidase B-like protein"
                     /protein_id="BAA90475.1"
                     /db_xref="GI:6855464"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQTGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHNRMWRKNRSFYANNHCIGTDL
                     NRNFASKHWCEEGASSSSCSETYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYS
                     QHIVFPYSYTRSKSKDHEELSLVASEAVRAIEKISKNTRYTHGHGSETLYLAPGGGDD
                     WIYDLGIKYSFTSNPPVEKLLPLSLK"
     polyA_site      1546
                     /note="21 A nucleotides"
BASE COUNT      453 a    323 c    297 g    473 t
ORIGIN

Alignment Scores:
Pred. No.:              2.72e-129       Length:         1546
Score:                  1327.00         Matches:        244
Percent Similarity:     99.19%          Conservative:   0
Best Local Similarity:  99.19%          Mismatches:     2
Query Match:            99.18%          Indels:         0
DB:                     9               Gaps:           0

US-09-980-881A-4 (1-246) x AB011969 (1-1546)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 421

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 481

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db        482 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 541

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGGATTGACTGTGGAATTCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 601

Qy         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        602 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 661

Qy        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        662 GGAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 721

Qy        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 781

Qy        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 841

Qy        161 TyrSerGlnHisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 901

Qy        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACC 961

Qy        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1021

Qy        221 TrpIleTyrAspLeuGlyIleLysTyrSerPheThrSerAsnProProValGluLysLeu 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 TGGATCTATGATTTGGGCATCAAATATTCGTTTACATCAAACCCACCTGTAGAGAAGCTT 1081

Qy        241 LeuProLeuSerLeuLys 246
              ||||||||||||||||||
Db       1082 TTGCCGCTGTCTCTAAAA 1099

RESULT 2
AR232183
LOCUS       AR232183               1625 bp    DNA     linear   PAT 20-DEC-2002
DEFINITION  Sequence 1 from patent US 6455294.
ACCESSION   AR232183
VERSION     AR232183.1  GI:27274083
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1625)
  AUTHORS   Gan,W., Ketchum,K.A., Di Francesco,V. and Beasley,E.M.
  TITLE     Isolated human protease proteins, nucleic acid molecules encoding
            human protease proteins, and uses thereof
  JOURNAL   Patent: US 6455294-A 1 24-SEP-2002;
FEATURES             Location/Qualifiers
     source          1. .1625
                     /organism="unknown"
BASE COUNT      488 a    334 c    315 g    488 t
ORIGIN

Alignment Scores:
Pred. No.:              2.48e-125       Length:         1625
Score:                  1289.50         Matches:        244
Percent Similarity:     92.42%          Conservative:   0
Best Local Similarity:  92.42%          Mismatches:     2
Query Match:            96.38%          Indels:         18
DB:                     6               Gaps:           1

US-09-980-881A-4 (1-246) x AR232183 (1-1625)
```

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 418

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        419 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 478

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db        479 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 538

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 598

Qy         81 IleGlyHisAsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIle 100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        599 ATAGGCCATAATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATC 658

Qy        101 GlyThrAspLeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSer 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        659 GGAACAGACCTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGT 718

Qy        121 SerSerCysSerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAla 140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        719 TCCTCATGCTCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCA 778

Qy        141 ValAlaSerPheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSer 160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        779 GTGGCTAGTTTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCA 838

Qy        161 TyrSerGlnHisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGlu 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TACTCCCAGCATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAG 898

Qy        181 GluLeuSerLeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThr 200
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||
Db        899 GAACTGTCTCTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAATTAGTAAAAATACC 958

Qy        201 ArgTyrThrHisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAsp 220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        959 AGGTATACACATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGAT 1018

Qy        221 TrpIleTyrAspLeuGlyIleLysTyrSerPhe-------------------------- 231
              |||||||||||||||||||||||||||||||
Db       1019 TGGATCTATGATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACA 1078

Qy        232 -------------------------ThrSerAsnProProValGluLysLeuLeuProLe 243
                                        ||||||||||||||||||||||||||||||||||
Db       1079 TACGGATTCTTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCT 1138

Qy        243 uSerLeuLys 246
              ||||||||||
Db       1139 GTCTCTAAAA 1148
```

```
RESULT 3
AR086324
LOCUS       AR086324               1272 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 1 from patent US 5985562.
ACCESSION   AR086324
VERSION     AR086324.1  GI:10013090
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.John. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk associated with plasma
            carboxypeptidase B polymorphisms
  JOURNAL   Patent: US 5985562-A 1 16-NOV-1999;
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="unknown"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN
```

Alignment Scores:

Pred. No.:	1.24e-123	Length:	1272
Score:	1272.00	Matches:	246
Percent Similarity:	81.73%	Conservative:	0
Best Local Similarity:	81.73%	Mismatches:	0
Query Match:	95.07%	Indels:	55
DB:	6	Gaps:	2

US-09-980-881A-4 (1-246) x AR086324 (1-1272)

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 642

Qy         83 ---------------------------------------------------------- 83

Db        643 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 882

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1062

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1182

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       1183 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
              |
Db       1243 A 1243

RESULT 4
BD106653
LOCUS       BD106653                1272 bp    DNA     linear   PAT 18-SEP-2002
DEFINITION  Method of detecting thrombotic disease risk.
ACCESSION   BD106653
VERSION     BD106653.1  GI:23201471
KEYWORDS    JP 2002502253-A/1.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Morser,M.J. and Nagashima,M.
  TITLE     Method of detecting thrombotic disease risk
  JOURNAL   Patent: JP 2002502253-A 1 22-JAN-2002;
            SCHERING AG
COMMENT     PN   JP 2002502253-A/1
            PD   22-JAN-2002
            PF   02-JUN-1998 JP 1999501465
            PR   03-JUN-1997 US    08/869057
            PI   MICHAEL JOHN MORSER,MARIKO NAGASHIMA
            PC   C12Q1/68
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32630"
BASE COUNT      375 a    269 c    271 g    357 t
ORIGIN

Alignment Scores:
Pred. No.:              1.24e-123      Length:         1272
Score:                  1272.00        Matches:        246
Percent Similarity:     81.73%         Conservative:   0
Best Local Similarity:  81.73%         Mismatches:     0
Query Match:            95.07%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-4 (1-246) x BD106653 (1-1272)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 642

Qy         83 ---------------------------------------------------------- 83

Db        643 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 882

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1062

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1182

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1183 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
              |
Db       1243 A 1243

RESULT 5
AX409750
LOCUS       AX409750                1728 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 2397 from Patent WO0229103.
ACCESSION   AX409750
VERSION     AX409750.1  GI:21442455
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 2397 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. M75106"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN

Alignment Scores:
Pred. No.:              1.84e-123      Length:         1728
Score:                  1272.00        Matches:        246
Percent Similarity:     81.73%         Conservative:   0
Best Local Similarity:  81.73%         Mismatches:     0
Query Match:            95.07%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-4 (1-246) x AX409750 (1-1728)
```

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 421

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 481

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 541

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 601

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        602 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 661

Qy         83 ---------------------------------------------------------- 83

Db        662 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 721

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 781

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 841

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 901

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 961

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1021

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1141

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1142 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1201

Qy        232 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                            ||||||||||||||||||||||||||||||||||||||||||||||
Db       1202 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1261

Qy        246 s 246
              |
Db       1262 A 1262
```

```
RESULT 6
AX706330
LOCUS       AX706330               1728 bp    DNA     linear   PAT 04-APR-2003
DEFINITION  Sequence 1 from Patent WO03014391.
ACCESSION   AX706330
VERSION     AX706330.1  GI:29562754
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Morten,J.E. and Jackson,K.N.
  TITLE     Method
  JOURNAL   Patent: WO 03014391-A 1 20-FEB-2003;
            AstraZeneca AB (SE)
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                1.84e-123       Length:          1728
Score:                    1272.00         Matches:         246
Percent Similarity:       81.73%          Conservative:    0
Best Local Similarity:    81.73%          Mismatches:      0
Query Match:              95.07%          Indels:          55
DB:                       6               Gaps:            2

US-09-980-881A-4 (1-246) x AX706330 (1-1728)
```

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 421

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 481

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 541

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 601

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        602 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 661

Qy         83 ---------------------------------------------------------- 83

Db        662 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 721

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 781

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 841

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 901

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 961

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1021

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081
```

```
Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1141

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1142 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1201

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1202 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1261

Qy        246 s 246
              |
Db       1262 A 1262
```

```
RESULT 7
HUMPCPBX
LOCUS       HUMPCPBX                1728 bp    mRNA    linear   PRI 07-JAN-1995
DEFINITION  Human prepro-plasma carboxypeptidase B mRNA, complete cds.
ACCESSION   M75106
VERSION     M75106.1  GI:189686
KEYWORDS    plasma carboxypeptidase.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1728)
  AUTHORS   Eaton,D.L., Malloy,B.E., Tsai,S.P., Henzel,W. and Drayna,D.
  TITLE     Isolation, molecular cloning, and partial characterization of a
            novel carboxypeptidase B from human plasma
  JOURNAL   J. Biol. Chem. 266 (32), 21833-21838 (1991)
  MEDLINE   92042093
   PUBMED   1939207
COMMENT     Original source text: Homo sapiens liver cDNA to mRNA.
FEATURES             Location/Qualifiers
     source          1. .1728
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /tissue_type="liver"
     gene            1. .1728
                     /gene="pCPB"
     CDS             20. .1291
                     /gene="pCPB"
                     /note="prepro-plasma carboxypeptidase B"
                     /codon_start=1
                     /protein_id="AAA60042.1"
                     /db_xref="GI:189687"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQTAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
     mat_peptide     86. .1288
                     /gene="pCPB"
                     /product="plasma carboxypeptidase B"
BASE COUNT      518 a    354 c    338 g    518 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.84e-123       Length:         1728
Score:                  1272.00         Matches:        246
Percent Similarity:     81.73%          Conservative:   0
Best Local Similarity:  81.73%          Mismatches:     0
Query Match:            95.07%          Indels:         55
DB:                     9               Gaps:           2

US-09-980-881A-4 (1-246) x HUMPCPBX (1-1728)
```

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        362 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 421

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        422 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 481

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        482 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 541

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        542 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 601

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||||
Db        602 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 661

Qy         83 ---------------------------------------------------------- 83

Db        662 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 721

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        722 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 781

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        782 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 841

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        842 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 901

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        902 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 961

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        962 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1021

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1022 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1081

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1082 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1141

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1142 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1201

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||||
Db       1202 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1261

Qy        246 s 246
              |
Db       1262 A 1262
```

```
RESULT 8
I16100
LOCUS       I16100                  1749 bp    DNA     linear   PAT 03-APR-1996
DEFINITION  Sequence 2 from patent US 5474901.
ACCESSION   I16100
VERSION     I16100.1  GI:1251008
KEYWORDS    .
SOURCE      Unknown.
```

```
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Antibodies to human carboxypeptidase B and methods of use thereof
  JOURNAL   Patent: US 5474901-A 2 12-DEC-1995;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:              1.87e-123      Length:         1749
Score:                  1272.00        Matches:        246
Percent Similarity:     81.73%         Conservative:   0
Best Local Similarity:  81.73%         Mismatches:     0
Query Match:            95.07%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-4 (1-246) x I16100 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ---------------------------------------------------------- 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                              ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
              |
Db       1283 A 1283

RESULT 9
I33526
LOCUS       I33526                  1749 bp    DNA     linear   PAT 06-FEB-1997
DEFINITION  Sequence 2 from patent US 5593674.
ACCESSION   I33526
VERSION     I33526.1  GI:1824317
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 1749)
  AUTHORS   Drayna,D.T. and Eaton,D.L.
  TITLE     Plasma carboxypeptidase
  JOURNAL   Patent: US 5593674-A 2 14-JAN-1997;
FEATURES             Location/Qualifiers
     source          1. .1749
                     /organism="unknown"
BASE COUNT      521 a    361 c    342 g    525 t
ORIGIN

Alignment Scores:
Pred. No.:              1.87e-123      Length:         1749
Score:                  1272.00        Matches:        246
Percent Similarity:     81.73%         Conservative:   0
Best Local Similarity:  81.73%         Mismatches:     0
Query Match:            95.07%         Indels:         55
DB:                     6              Gaps:           2

US-09-980-881A-4 (1-246) x I33526 (1-1749)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 442

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 502

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAACAGCCAAAAATGCCATA 562

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        563 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 622

Qy         81 IleGlyHis---------------------------------------------------- 83
              |||||||||
Db        623 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 682

Qy         83 ---------------------------------------------------------- 83

Db        683 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGACGGTTATGACTACTCATGGAAAAAG 742

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        743 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 802

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        803 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 862

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        863 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 922

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        923 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 982

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        983 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1042

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1043 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1102

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1103 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1162

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1163 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1222

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||||||||||||||||||||||||||||||||||
Db       1223 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1282

Qy        246 s 246
              |
Db       1283 A 1283
```

```
RESULT 10
BT006936
LOCUS       BT006936               1272 bp    mRNA    linear   PRI 13-MAY-2003
DEFINITION  Homo sapiens carboxypeptidase B2 (plasma, carboxypeptidase U) mRNA,
            complete cds.
ACCESSION   BT006936
VERSION     BT006936.1  GI:30582710
KEYWORDS    FLI_CDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GH00156X1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .1272
                     /codon_start=1
                     /product="carboxypeptidase B2 (plasma, carboxypeptidase
                     U)"
                     /protein_id="AAP35582.1"
                     /db_xref="GI:30582711"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT      373 a    268 c    273 g    358 t
ORIGIN

Alignment Scores:
Pred. No.:              4.14e-123       Length:         1272
Score:                  1267.00         Matches:        245
Percent Similarity:     81.40%          Conservative:   0
Best Local Similarity:  81.40%          Mismatches:     1
Query Match:            94.69%          Indels:         55
DB:                     9               Gaps:           2

US-09-980-881A-4 (1-246) x BT006936 (1-1272)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis------------------------------------------------- 83
              ||||||||
Db        583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 642

Qy         83 ---------------------------------------------------------- 83

Db        643 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 882

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942
```

```
Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1062

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1182

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||
Db       1183 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
              |
Db       1243 A 1243
```

```
RESULT 11
BT007990
LOCUS       BT007990               1272 bp    mRNA    linear   SYN 13-MAY-2003
DEFINITION  Synthetic construct Homo sapiens carboxypeptidase B2 (plasma,
            carboxypeptidase U) mRNA, partial cds.
ACCESSION   BT007990
VERSION     BT007990.1  GI:30584818
KEYWORDS    FLI_CDNA.
SOURCE      synthetic construct
  ORGANISM  synthetic construct
            artificial sequences.
REFERENCE   1  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Cloning of human full-length CDSs in BD Creator(TM) System Donor
            vector
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1272)
  AUTHORS   Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S.,
            Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y.,
            Phelan,M. and Farmer,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow
            Circle, Palo Alto, CA 94303, USA
COMMENT     This CDS clone is a part of a collection of human full length
            expression clones generated by BD Biosciences Clontech and the
            Harvard Institute of Proteomics. Each CDS has been cloned in two
            forms: with and without stop-codon (to allow fusion with C-terminal
            tag). The CDS has been  directionally cloned using BD In-Fusion(TM)
            cloning system between the SalI  and HindIII sites of the pDNR-DUAL
            vector. Additional sequences in the clone:  'ACC' after SalI site
            and before 'ATG' to provide Kozak consensus sequence; 'GG' after
            last codon and before HindIII site to maintain reading frame.
            Clone distribution: http://bioinfo.clontech.com/orfclones.
FEATURES             Location/Qualifiers
     source          1. .1272
                     /organism="synthetic construct"
                     /mol_type="mRNA"
                     /db_xref="taxon:32630"
                     /clone="GH00156L1.0"
                     /clone_lib="BD Creator(TM) CDS Library derived from MGC
                     collection"
                     /lab_host="DH5alpha T1 resistant"
                     /note="Vector: pDNR-Dual"
     CDS             1. .>1272
                     /note="Mutations: 1271:Stop->Leu"
                     /codon_start=1
                     /transl_table=11
                     /product="Homo sapiens carboxypeptidase B2 (plasma,
                     carboxypeptidase U)"
                     /protein_id="AAP36662.1"
                     /db_xref="GI:30584819"
                     /translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
                     LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
                     QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
                     LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
                     VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
                     TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
                     LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
                     GFLLPERYIKPTCREAFAAVSKIAWHVIRNVL"
BASE COUNT      372 a    268 c    273 g    359 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:                4.14e-123      Length:          1272
Score:                    1267.00        Matches:         245
Percent Similarity:       81.40%         Conservative:    0
Best Local Similarity:    81.40%         Mismatches:      1
Query Match:              94.69%         Indels:          55
DB:                       12             Gaps:            2

US-09-980-881A-4 (1-246) x BT007990 (1-1272)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        343 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 402

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        403 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 462

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||  ||||||||||||||||
Db        463 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 522

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        523 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 582

Qy         81 IleGlyHis------------------------------------------------- 83
              ||||||||
Db        583 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 642

Qy         83 ------------------------------------------------------------ 83

Db        643 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAG 702

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        703 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 762

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        763 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 822

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        823 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 882

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        883 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 942

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        943 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1002

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1003 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1062
```

```
Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1063 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1122

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1123 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1182

Qy        232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                               ||||||||||||||||||||||||||||||||||||||||||
Db       1183 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTTGCCGCTGTCTCTAAA 1242

Qy        246 s 246
              |
Db       1243 A 1243
```

RESULT 12
BC007057
LOCUS BC007057 1715 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, carboxypeptidase B2 (plasma), clone MGC:12495 IMAGE:3934520, mRNA, complete cds.
ACCESSION BC007057
VERSION BC007057.1 GI:13937896
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1715)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (30-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 16 Row: f Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4503004.

FEATURES Location/Qualifiers
source 1. .1715
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="LocusID:1361"
/db_xref="taxon:9606"
/clone="MGC:12495 IMAGE:3934520"
/tissue_type="Skeletal Muscle"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B"
/note="Vector: pDNR-LIB"
CDS 25. .1296
/codon_start=1
/product="carboxypeptidase B2 (plasma)"
/protein_id="AAH07057.1"
/db_xref="GI:13937897"
/translation="MKLCSLAVLVPIVLFCEQHVFAFQSGQVLAALPRTSRQVQVLQN
LTTTYEIVLWQPVTADLIVKKKQVHFFVNASDVDNVKAHLNVSGIPCSVLLADVEDLI
QQQISNDTVSPRASASYYEQYHSLNEIYSWIEFITERHPDMLTKIHIGSSFEKYPLYV
LKVSGKEQAAKNAIWIDCGIHAREWISPAFCLWFIGHITQFYGIIGQYTNLLRLVDFY
VMPVVNVDGYDYSWKKNRMWRKNRSFYANNHCIGTDLNRNFASKHWCEEGASSSSCSE
TYCGLYPESEPEVKAVASFLRRNINQIKAYISMHSYSQHIVFPYSYTRSKSKDHEELS
LVASEAVRAIEKTSKNTRYTHGHGSETLYLAPGGGDDWIYDLGIKYSFTIELRDTGTY
GFLLPERYIKPTCREAFAAVSKIAWHVIRNV"
BASE COUNT 518 a 348 c 341 g 508 t
ORIGIN

Alignment Scores:

Pred. No.:	6.09e-123	Length:	1715
Score:	1267.00	Matches:	245
Percent Similarity:	81.40%	Conservative:	0
Best Local Similarity:	81.40%	Mismatches:	1
Query Match:	94.69%	Indels:	55
DB:	9	Gaps:	2

US-09-980-881A-4 (1-246) x BC007057 (1-1715)

```
Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        367 GCCTCCGCATCGTACTATGAACAGTATCACTCACTAAATGAAATCTATTCTTGGATAGAA 426

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        427 TTTATAACTGAGAGGCATCCTGATATGCTTACAAAAATCCACATTGGATCCTCATTTGAG 486

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db        487 AAGTACCCACTCTATGTTTTAAAGGTTTCTGGAAAAGAACAAGCAGCCAAAAATGCCATA 546

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        547 TGGATTGACTGTGGAATCCATGCCAGAGAATGGATCTCTCCTGCTTTCTGCTTGTGGTTC 606

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||||
Db        607 ATAGGCCATATAACTCAATTCTATGGGATAATAGGGCAATATACCAATCTCCTGAGGCTT 666

Qy         83 ---------------------------------------------------------- 83

Db        667 GTGGATTTCTATGTTATGCCGGTGGTTAATGTGGATGGTTATGACTACTCATGGAAAAAG 726

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        727 AATCGAATGTGGAGAAAGAACCGTTCTTTCTATGCGAACAATCATTGCATCGGAACAGAC 786

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        787 CTGAATAGGAACTTTGCTTCCAAACACTGGTGTGAGGAAGGTGCATCCAGTTCCTCATGC 846

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        847 TCGGAAACCTACTGTGGACTTTATCCTGAGTCAGAACCAGAAGTGAAGGCAGTGGCTAGT 906

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        907 TTCTTGAGAAGAAATATCAACCAGATTAAAGCATACATCAGCATGCATTCATACTCCCAG 966

Qy        164 HisIleValProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        967 CATATAGTGTTTCCATATTCCTATACACGAAGTAAAAGCAAAGACCATGAGGAACTGTCT 1026

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1027 CTAGTAGCCAGTGAAGCAGTTCGTGCTATTGAGAAAACTAGTAAAAATACCAGGTATACA 1086

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1087 CATGGCCATGGCTCAGAAACCTTATACCTAGCTCCTGGAGGTGGGGACGATTGGATCTAT 1146

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1147 GATTTGGGCATCAAATATTCGTTTACAATTGAACTTCGAGATACGGGCACATACGGATTC 1206
```

```
Qy       232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||||||||||||||||||||||||||||||||||
Db      1207 TTGCTGCCGGAGCGTTACATCAAACCCACCTGTAGAGAAGCTTTGCCGCTGTCTCTAAA 1266

Qy       246 s 246
             |
Db      1267 A 1267
```

RESULT 13
AB021968
LOCUS AB021968 1421 bp mRNA linear ROD 27-JUL-2000
DEFINITION Mus musculus mRNA for carboxypeptidase R, complete cds.
ACCESSION AB021968
VERSION AB021968.1 GI:9558447
KEYWORDS carboxypeptidase R.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Sato,T., Miwa,T., Akatsu,H., Matsukawa,N., Obata,K., Okada,N., Campbell,W. and Okada,H.
TITLE Pro-carboxypeptidase R is an acute phase protein in the mouse, whereas carboxypeptidase N is not
JOURNAL J. Immunol. 165 (2), 1053-1058 (2000)
MEDLINE 20341711
PUBMED 10878383
REFERENCE 2 (bases 1 to 1421)
AUTHORS Sato,T.
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1998) Tomoo Sato, Fukushimura Hospital, Choju Medical Institute; 19-14, Azayamanaka, Noyori-cho, Toyohashi, Aichi 441-8124, Japan (E-mail:tomoo@tcp-ip.or.jp, Tel:81-532-46-7511, Fax:81-532-46-4899)
FEATURES Location/Qualifiers
source 1. .1421
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
CDS 17. .1285
/codon_start=1
/product="carboxypeptidase R"
/protein_id="BAB03402.1"
/db_xref="GI:9558448"
/translation="MKLHGLGILVAIILYEQHGFAPQSGQVLSALPRTSRQVQLLQNL
TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
FLLPERYIKPTCAEALAAISKIVWHVIRNT"
polyA_site 1421
/note="11 a nucleotides"
BASE COUNT 418 a 320 c 305 g 378 t
ORIGIN

Alignment Scores:

Pred. No.:	8.64e-107	Length:	1421
Score:	1112.00	Matches:	217
Percent Similarity:	75.75%	Conservative:	11
Best Local Similarity:	72.09%	Mismatches:	18
Query Match:	83.11%	Indels:	55
DB:	10	Gaps:	2

US-09-980-881A-4 (1-246) x AB021968 (1-1421)

```
Qy         1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       356 GCCTCCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAA 415

Qy        21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 ||||||||:::|||||||||||||||||   ||||||:::||||||||||||||||||
Db       416 GTCATAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAG 475

Qy        41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       476 AAGTACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATC 535

Qy        61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       536 TGGATCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTC 595

Qy        81 IleGlyHis---------------------------------------------------- 83
             ||||||:::
Db       596 ATAGGCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCAC 655

Qy        83 ---------------------------------------------------------- 83

Db       656 GTGGATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAG 715

Qy        84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
             ||||||||||||||||||||||||||   :::   ||||||   |||:::||||||||
Db       716 AATCGAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAC 775

Qy       104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
             ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db       776 CTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 835

Qy       124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       836 TCTGAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 895

Qy       144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
             ||||||||||||||||||:::   ||||||||||||||||||||||||||||||||||||
Db       896 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAA 955

Qy       164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
               |||:::||||||||||||||   ||||||||||||||||||||||||||||||||||||
Db       956 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCT 1015

Qy       184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
             ||||||||||||||||||||||||||||||||||||      :::||||||||||||||||
Db      1016 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACA 1075

Qy       204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
             ||||||   ||||||||||:::||||||||||||||||||||   ||||||||||||||||
Db      1076 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 1135

Qy       224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
             |||||||||||||||||||||||
Db      1136 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 1195

Qy       232 ----------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                             ||||||||||||   |||:::||||||   |||   |||||||||
Db      1196 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 1255

Qy       246 s 246
             |
Db      1256 A 1256
```

RESULT 14
AF164524
LOCUS AF164524 1490 bp mRNA linear ROD 05-APR-2000
DEFINITION Mus musculus thrombin-activatable fibrinolysis inhibitor (Tafi) mRNA, complete cds.
ACCESSION AF164524
VERSION AF164524.1 GI:7416966
KEYWORDS .
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1490)
AUTHORS Marx,P.F., Wagenaar,G.T., Reijerkerk,A., Tiekstra,M.J., van

```
          Rossum,A.G., Gebbink,M.F. and Meijers,J.C.
  TITLE   Characterization of mouse thrombin-activatable fibrinolysis
          inhibitor
  JOURNAL Thromb. Haemost. 83 (2), 297-303 (2000)
  MEDLINE 20201996
   PUBMED 10739389
REFERENCE 2  (bases 1 to 1490)
  AUTHORS Marx,P.F., Wagenaar,G.T.M., van Rossum,A.G.S.H. and Meijers,J.C.M.
  TITLE   Direct Submission
  JOURNAL Submitted (01-JUL-1999) Haematology, UMCU, Heidelberglaan 100,
          Utrecht 3584 CX, The Netherlands
FEATURES             Location/Qualifiers
     source          1. .1490
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:477035"
     gene            1. .1490
                     /gene="Tafi"
     CDS             84. .1352
                     /gene="Tafi"
                     /function="carboxypeptidase"
                     /codon_start=1
                     /product="thrombin-activatable fibrinolysis inhibitor"
                     /protein_id="AAF62385.1"
                     /db_xref="GI:7416967"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVITEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIESINKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      438 a    337 c    322 g    393 t
ORIGIN

Alignment Scores:
Pred. No.:              9.18e-107       Length:         1490
Score:                  1112.00         Matches:        217
Percent Similarity:     75.75%          Conservative:   11
Best Local Similarity:  72.09%          Mismatches:     18
Query Match:            83.11%          Indels:         55
DB:                     10              Gaps:           2

US-09-980-881A-4 (1-246) x AF164524 (1-1490)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        423 GCCTCCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAA 482

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 |||||||||:::|||||||||||||||    ||||||:::|||||||||||||||||
Db        483 GTCATAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAG 542

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||      ||||||||||||
Db        543 AAGTACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATC 602

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        603 TGGATCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTC 662

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||:::
Db        663 ATAGGCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCAC 722

Qy         83 ------------------------------------------------------------ 83

Db        723 GTGGATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAG 782

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              |||||||||||||||||||||||||||   :::   ||||||   |||:::||||||||||

Db        783 AATCGAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAC 842

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db        843 CTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 902

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        903 TCTGAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 962

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||:::   ||||||||||||||||||||||||||||||||||||
Db        963 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAA 1022

Qy        164 HisIleValProProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||
Db       1023 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCT 1082

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||||      :::|||||||||||||||||||
Db       1083 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGAAAGTATTAATAAAAACACCAGGTACACA 1142

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||   |||||||||:::|||||||||||||||||||||   ||||||||||||||
Db       1143 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 1202

Qy        224 AspLeuGlyIleLysTyrSerPhe------------------------------------ 231
              |||||||||||||||||||||||
Db       1203 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 1262

Qy        232 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                            ||||||||||||   |||:::||||||   |||   ||||||||
Db       1263 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 1322

Qy        246 s 246
              |
Db       1323 A 1323

RESULT 15
AF186188
LOCUS       AF186188               1430 bp    mRNA    linear   ROD 01-OCT-1999
DEFINITION  Mus musculus carboxypeptidase U mRNA, complete cds.
ACCESSION   AF186188
VERSION     AF186188.1  GI:6003651
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Isolation and characterization of mouse liver carboxypeptidase B
            gene
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1430)
  AUTHORS   He,Y.C. and Broze,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-SEP-1999) Division of Hematology, Washington
            University School of Medicine, 216 S. Kingshighway, St. Louis, MO
            63110, USA
FEATURES             Location/Qualifiers
     source          1. .1430
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="B6/CBAF1J"
                     /db_xref="taxon:10090"
     CDS             20. .1288
                     /note="pro-carboxypeptidase-B; thrombin-activatable
                     fibrinolysis inhibitor (TAFI); pCPU; pCPB; binds
                     plasminogen"
                     /codon_start=1
                     /product="carboxypeptidase U"
                     /protein_id="AAF00528.1"
```

```
                     /db_xref="GI:6003652"
                     /translation="MKLHGLGILVAIILYEQHGFAFQSGQVLSALPRTSRQVQLLQNL
                     TTTYEVVLWQPVTAEFIEKKKEVHFFVNASDVDSVKAHLNVSRIPFNVLMNNVEDLIE
                     QQTFNDTVSPRASASYYEQYHSLNEIYSWIEVLTEQHPDMLQKIYIGSSFEKYPLYVL
                     KVSGKEQRIKNAIWIDCGIHAREWISPAFCLWFIGYVTQFHGKENLYTRLLRHVDFYI
                     MPVMNVDGYDYTWKKNRMWRKNRSAHKNNRCVGTDLNRNFASKHWCEKGASSSSCSET
                     YCGLYPESEPEVKAVADFLRRNIDHIKAYISMHSYSQQILFPYSYNRSKSKDHEELSL
                     VASEAVRAIGSFNKNTRYTHGSGSESLYLAPGGSDDWIYDLGIKYSFTIELRDTGRYG
                     FLLPERYIKPTCAEALAAISKIVWHVIRNT"
BASE COUNT      421 a    320 c    310 g    379 t
ORIGIN

Alignment Scores:
Pred. No.:              9.74e-106       Length:         1430
Score:                  1102.00         Matches:        215
Percent Similarity:     75.42%          Conservative:   12
Best Local Similarity:  71.43%          Mismatches:     19
Query Match:            82.36%          Indels:         55
DB:                     10              Gaps:           2

US-09-980-881A-4 (1-246) x AF186188 (1-1430)

Qy          1 AlaSerAlaSerTyrTyrGluGlnTyrHisSerLeuAsnGluIleTyrSerTrpIleGlu 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        359 GCCTCCGCTTCATACTATGAGCAGTATCACTCGCTAAATGAAATCTATTCCTGGATAGAA 418

Qy         21 PheIleThrGluArgHisProAspMetLeuThrLysIleHisIleGlySerSerPheGlu 40
                 :::||||||:::|||||||||||||||||    ||||||:::||||||||||||||||||
Db        419 GTCTTAACTGAACAGCATCCTGACATGCTCCAGAAAATCTACATCGGATCATCATTCGAG 478

Qy         41 LysTyrProLeuTyrValLeuLysValSerGlyLysGluGlnThrAlaLysAsnAlaIle 60
              ||||||||||||||||||||||||||||||||||||||||      ||||||||||||
Db        479 AAGTACCCACTTTATGTTTTAAAGGTCTCAGGAAAGGAACAAAGAATCAAAAATGCCATC 538

Qy         61 TrpIleAspCysGlyIleHisAlaArgGluTrpIleSerProAlaPheCysLeuTrpPhe 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        539 TGGATCGACTGTGGAATCCATGCCAGAGAATGGATTTCACCTGCTTTCTGTTTGTGGTTC 598

Qy         81 IleGlyHis---------------------------------------------------- 83
              ||||||:::
Db        599 ATAGGCTACGTGACACAATTCCATGGGAAAGAAAATCTGTATACCAGACTTCTGAGGCAC 658

Qy         83 ------------------------------------------------------------ 83

Db        659 GTGGATTTCTACATCATGCCCGTGATGAACGTGGATGGCTATGACTACACGTGGAAAAAG 718

Qy         84 AsnArgMetTrpArgLysAsnArgSerPheTyrAlaAsnAsnHisCysIleGlyThrAsp 103
              ||||||||||||||||||||||||||    :::   ||||||   |||:::||||||||||
Db        719 AATCGAATGTGGAGGAAGAACCGCTCTGCTCACAAGAACAACCGCTGCGTGGGCACAGAC 778

Qy        104 LeuAsnArgAsnPheAlaSerLysHisTrpCysGluGluGlyAlaSerSerSerSerCys 123
              ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db        779 CTGAACAGGAACTTCGCTTCCAAACACTGGTGTGAGAAAGGTGCGTCAAGTTCCTCCTGC 838

Qy        124 SerGluThrTyrCysGlyLeuTyrProGluSerGluProGluValLysAlaValAlaSer 143
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        839 TCTGAAACCTACTGTGGACTTTATCCTGAGTCTGAGCCAGAGGTGAAGGCAGTGGCTGAC 898

Qy        144 PheLeuArgArgAsnIleAsnGlnIleLysAlaTyrIleSerMetHisSerTyrSerGln 163
              ||||||||||||||||||:::   ||||||||||||||||||||||||||||||||||||
Db        899 TTCTTGAGAAGAAATATCGACCACATTAAAGCTTACATCAGTATGCACTCATACTCCCAA 958

Qy        164 HisIleValPheProTyrSerTyrThrArgSerLysSerLysAspHisGluGluLeuSer 183
                 |||:::|||||||||||||||   ||||||||||||||||||||||||||||||||
Db        959 CAAATACTGTTTCCCTATTCCTATAACAGAAGCAAAAGCAAGGACCACGAAGAACTGTCT 1018

Qy        184 LeuValAlaSerGluAlaValArgAlaIleGluLysThrSerLysAsnThrArgTyrThr 203
              ||||||||||||||||||||||||||||||         :::||||||||||||||||||
Db       1019 CTAGTGGCCAGCGAAGCAGTTCGTGCAATTGGAAGTTTTAATAAAAACACCAGGTACACA 1078

Qy        204 HisGlyHisGlySerGluThrLeuTyrLeuAlaProGlyGlyGlyAspAspTrpIleTyr 223
              ||||||   |||||||||:::||||||||||||||||||||   ||||||||||||||||
Db       1079 CACGGCAGTGGCTCAGAAAGTTTATATCTAGCTCCTGGAGGTTCTGACGATTGGATCTAT 1138

Qy        224 AspLeuGlyIleLysTyrSerPhe---------------------------------- 231
              |||||||||||||||||||||||
Db       1139 GATTTGGGCATCAAATATTCGTTTACAATTGAGCTCCGAGATACAGGCAGATACGGATTC 1198

Qy        232 ---------------ThrSerAsnProProValGluLysLeuLeuProLeuSerLeuLy 246
                            ||||||||||||   |||:::||||||   |||   ||||||||
Db       1199 TTGCTGCCTGAGAGATACATCAAACCCACTTGTGCAGAAGCTTTGGCCGCCATCTCTAAA 1258

Qy        246 s 246
              |
Db       1259 A 1259
```

Search completed: January 2, 2004, 00:16:40
Job time : 3811.88 secs

This Page Blank (uspto)